BIODIVERSITAS
Volume 21, Number 8, August 2020
Pages: 3879-3885

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d210858

An in silico approach for evaluation of *rbc*L and *mat*K loci for DNA barcoding of Cucurbitaceae family

VIET THE HO♥, MINH PHUONG NGUYEN
Faculty of Biotechnology, Ho Chi Minh City University of Food Industry. 140 Le Trong Tan, Tan Phu District, Ho Chi Minh City, Vietnam.
Tel./fax.: +84-933-25523, ♥email: thehv@hufi.edu.vn

Manuscript received: 25 June 2020. Revision accepted: 30 July 2020.

Abstract. *Ho VT, Nguyen MP. 2020. An in silico approach for evaluation of* rbcL *and* matK *loci for DNA barcoding of Cucurbitaceae family. Biodiversitas 21: 3879-3885.* DNA barcodes have been used intensively to discriminate different species in Cucurbitaceae family. The main of this study is to evaluate the effectiveness of *rbc*L and *mat*K loci for 16 species of Cucurbitaceae family by using in silico approach. For analysis, sequences were firstly retrieved from NCBI and then calculated for sequence parameters. Sequences were then aligned and constructed phylogenetic try and examined for species resolution ability. The obtained data show the variability of resolving capacity among species. *rbc*L region is suitable for distinguishing five species namely *S. edule, M. cochinchinensis, L. aegyptiaca, C. melo,* and *C. pepo*, whereas *mat*K locus is more proper for different five species consisting of *M. balsamina, M. cochinchinensis, M. charantia, S. edule,* and *C. sativus*. The resolving power is improved sharply by analyzing the *rbc*L + *mat*K combination with up to nine species consisting of *C. lanatus, B. hispida, C. melo, C. sativus, C. pepo, C. agryrosperma, L. aegyptiaca, S. edule,* and *M. cochinchinensis*. Therefore, the integration of *rbc*L and *mat*K loci may improve the competence of assessing genetic relatedness at species level of members in Cucurbitaceae family. The obtained information could be important for choosing proper DNA barcode loci for phylogenetic study of this crop family.

Keywords: Cucurbitaceae**,** DNA barcode, *mat*K, *rbc*L

INTRODUCTION

Cucurbitaceae is a big crop family of crop plant. There are more than 120 genera with 800 species have been reported (Welbaum, 2014) or up to 1000 species (Chomicki et al. 2019). But only 10 species are considered economic importance and cultivated worldwide and another 23 species are important in certain areas (Chomicki et al. 2019). These species are cultivated mostly for their fruit which is consumed in different ways from ripe such as pumpkin, unripe such as zucchini, raw such watermelon, cooked such as squash or even pickle such as baby cucumber. Because of its economic importance, several studies have been carried out to increase quality and yield of this plant. However, the classification based on morphology such as pollen structure, seed coat character, flower characters shows difficulty for identification. Although morphological classification is easy to perform and to carry out on the field with low cost, this method shows several limitations such as limited number, complex inheritance pattern, and vulnerable to changes of environment (Ahmedand and Mohamed, 2014).

Since 2003, DNA barcode has been used intensively for organism identification. DNA barcode is a short standardized DNA region used to identify organisms as specific as species level. This method possesses several advantages in the comparison to traditional methods such as high repeatability and stability, applicability to any developmental stages of organism, and ability to identify target organisms after cooked or processed. Various molecular markers have been used for DNA barcoding studies. In animal, mitochondrial cytochrome c oxidase subunits I (*COI*) was proposed and successfully applied to animal, birds, fish, insect, and nematode identification (Herbet et al. 2003; Huxley-Jones et al. 2012; Lijtmaer et al. 2012, Kiewnick et al. 2014; Knebelsberger et al. 2014). However, COI is not applicable to plant due to its slow evolution. Several studies have tried to investigate DNA barcode candidates for plant identification namely *trn*H, *psb*A, *rpo*C1, *rpo*B, *atp*F, *atp*H, *psb*K, and *psb*I. Consequently, massive data of DNA barcodes in a wide range of organisms, accumulated year by year despite its short history, is presently available on online databases ready to be utilized. Several recent papers have reviewed DNA barcoding in plants (Vijayan and Tsou, 2010; Hollingsworth et al. 2011). After that, many loci in the chloroplast genome have been tested for barcoding and ribulose-1, 5-biphosphate carboxylase (*rbc*L), and maturase K (*mat*K) have been proposed as preferred plant barcoding loci by consortium for the barcodes of life (CBOL, 2009). These two gene regions have also been used intensively for study in different species of Cucurbitaceae family (Kocyan et al. 2007; Kates et al. 2017; Kumar et al. 2020).

However, until now, the single and universal locus for DNA barcoding of plants remains debatable since some loci have just been efficient for some specific taxonomic groups and the species discrimination of these genes is variable among plant species. When studied *Aquilaria genus*, Thitikornpong and colleagues found that the *mat*K gene showed higher variation in the comparison to *rbc*L gene (Thitikornpong et al. 2018), a similar result has also found in phylogenetic analysis of *Dalbergia* (Li et al. 2017). *rbc*L and *mat*K also show variation in species resolution of different vascular plant species (Saarela et al.

2013), whereas, *rbc*L was shown more preferable in Teak, Black Rosewood, and Ben Teak (Fatima et al. 2019), liverwort (Hollingsworth et al. 2009). The availability of large public sequence databases may allow comparing multiple potential barcodes and their properties before performing studies by using in silico analysis. Applying this method, *mat*K gene was found to be effective in RNA splicing mechanisms (Mustafa et al. 2018). Similar in silico study of Fabaceae family conducted in Indian found that *mat*K gene is ideal for *Vigna*, *Cassia*, and *Crotalaria* species whilst *rbc*L is more efficient to *Sesbania* family (Sikdar et al. 2018). Therefore, the main aims of this in silico study were to evaluate the species resolution ability of *mat*K and *rbc*L loci in Cucurbitaceae family available on the National Center for Biotechnology Information (NCBI). This study may contribute to the use of proper molecular sequences as effective barcoding markers for identification of specific species of Cucurbitaceae family, serving for breeding, conservation, and diversity research of this crop plant.

MATERIALS AND METHODS

The DNA sequences of *rbc*L and *mat*K genes belonging to Cucurbitaceae family were retrieved from nucleotide database of NCBI (URL: http://www.ncbi.nlm.nih.gov). All of the sequences were subjected to critical evaluation and any low-quality sequences were removed. Criteria used to filter the sequences were (i) sequences are not 'unverified' without a species name (ii) contain <3% ambiguous base 'N' (Suesatpanit et al. 2017). The species details of each locus with their accession number are shown in Table 1.

For analyses, sequences of each species were saved in FASTA format. The average size was calculated. The sequences were subjected to Multiple Sequence Alignment using Clustal W by using MEGA 6 software (Tamura et al. 2013). Evolutionary divergence for each data set and pattern of nucleotide substitution was performed by the same software. For phylogenetic analysis, we used Neighbor-Joining tree method with 1000 bootstrap and presented as circular cladograms. In order to estimate species resolution for a given barcode locus, we considered the species were resolved if conspecific individuals are grouped into one monophyletic branch in the phylogenetic tree with well bootstrap support. In converse, if conspecific individuals are separated in paraphyletic branches, then the species is considered as identification failure (Sikdar et al. 2018).

RESULTS AND DISCUSSION

By using keyword "species+*rbc*L/matK" to find the sequences deposited in NCBI GenBank, after removal of unrealizable sequences as described by Suesatpanit and colleagues (2017), there are 125 and 170 DNA sequences of *rbc*L and *mat*K regions, respectively (Table 1).

In general, both *rbc*L and *mat*K are utilized equally in study different species of Cucurbitaceae family, except only *mat*K region was used for *M. balsamina* and no *rbc*L sequence was found for this species. The length of sequences is highly variable ranging from 522 bp to 1401 bp of *rcbL* and 512 bp to 1232 bp for *mat*K. *C. melo* shows highest sequence abundance for both regions but *mat*K sequences are at least two times higher than *rcbL* region suggesting the economic importance of this species. Due to the imbalance of sequence numbers among species, if any species showed higher than five sequences then only five random sequences were used for continuing analysis. Due to the high variability of sequences that could result in the inaccuracy of results, all positions containing gaps and missing data were eliminated. Finally, sequence length of *rbc*L and *mat*K regions used for analysis was 373 bp and 381 bp, respectively.

Table 1. List of plant species and number of retrieved sequences used in this study

Latin name	Genus	Common name	*rbc*L		*mat*K	
			Number of sequence for analysis*	Length range (bp)	Number of sequence for analysis*	Length range (bp)
Cucurbita pepo	*Cucurbita*	Pumpkin	5 (11)	552-1402	5 (10)	576-1149
Cucurbita maxima		Pumpkin	1 (1)	1401	2 (2)	802-1149
Cucurbita argyrosperma		Pumkin	5 (5)	1401	5 (5)	1149
Cucurbita moschata		Pumpkin	1 (1)	1401	1 (1)	1149
Lagenaria siceraria	*Lagenaria*	Calabash	5 (5)	514-1370	5 (8)	512-1055
Citrullus lanatus	*Citrullus*	Watermelon	5 (10)	549-1384	3 (3)	813-1151
Cucumis sativus	*Cucumis*	Cucumber	5 (18)	542-1428	5 (18)	576-1154
Cucumis melo		Muskmelon	5 (32)	703-1358	5 (71)	460-1157
Momordica charantia	*Momordica*	Bitter melon (domesticated plant)	5 (15)	522-1264	5 (23)	636-1232
Momordica balsamina		Bitter melon (wild plant)	-	-	5 (5)	767-1207
Momordica cochinchinensis		Gac	4 (4)	562-1355	5 (6)	740-1201
Luffa aegyptiaca	*Luffa*	Sponge gourd	5 (6)	570-1387	4 (4)	777-1157
Luffa acutangula		Sponge gourd	5 (11)	570-1423	5 (7)	776-1157
Benincasa hispida	*Benincasa*	Wax gourd	4 (4)	1305-1428	5 (6)	763-1081
Sechium edule	*Sechium*	Chayote	2 (2)	1286-1428	1 (1)	1130

Note: *: total number retrieved from NCBI GenBank of corresponding species is shown in parenthesis

Estimation of sequence divergence

The variation within and between species based on *rcbL* and *mat*K regions was calculated and the data representing by evolutionary divergence (p distance value) are shown in Table 2. Due to the only single *rbc*L sequence available of *C. moschata* and *mat*K of *C. moschata* and *S. edule*, the divergence value was not able to computed and named as "n/c". Based on p distance in Table 2, it suggests that there is a large variability of *rbc*L and *mat*K regions within a specific species of Cucurbitaceae family. Only *C. argyrosperma* shows the consistency of examining sequences with p-value as zero. There are two species present no variation in *rcbL* region namely *C. argyrosperma* and *C. pepo*, whereas up to four species show no variation in *mat*K region including *C. maxima, C. argyrosperma, C. lanatus*, and *L. aegyptiaca*. *M. charantia* and *B. hispida* show the highest variation in *rbc*L (0.324) and *mat*K (0.017) region, respectively.

The variability among *rbc*L and *mat*K regions is from 0 to 0.24 and 0.03, respectively (Tables 3 and 4). In *rbc*L region, *M. charantia* also shows a higher difference with those of other species which vary from 0.23 to 0.24 (*M. charantia* vs *B. hispida*). Whereas, *mat*K region of *B. hispida* and *S. edule* is shown the highest divergence which p distance ranging from 0.02 to 0.03. This value is significantly lower than previous data reported by Sikdar and colleagues when analyzing 46 *rbc*L sequences and 42 *mat*K sequences of 21 species in Fabaceae family (Sikdar et al. 2018). In 2012, a study in the effectiveness of *rbc*L and *mat*K regions for 490 vascular plant species, and *mat*K is more divergent than *rbc*L at both intraspecies and interspecies (Saarela et al. 2013).

Table 2. Estimates of average evolutionary Divergence of *rbc*L and *mat*K regions

Species	p distance of *rcb*L	SE	p distance of *mat*K	SE
C.pepo	0	0	0.001	0.001
C. maxima	n/c	n/c	0	0
C. argyrosperma	0	0	0	0
C. moschata	n/c	n/c	n/c	n/c
L. siceraria	0.002	0.002	0.005	0.003
C. lanatus	0.003	0.002	0	0.000
C. sativus	0.003	0.002	0.010	0.003
C. melo	0.001	0.001	0.009	0.003
M. charantia	0.324	0.015	0.001	0.001
M. balsamina	-	-	0.002	0.001
M. cochinchinensis	0.002	0.002	0.006	0.003
L. aegyptiaca	0.002	0.002	0	0
L. acutangula	0.007	0.003	0.002	0.002
B. hispida	0.001	0.001	0.017	0.004
S. edule	0.003	0.003	n/c	n/c

Note: SE: standard error; n/c: not-computed

Table 3. Estimates of evolutionary divergence in *rbc*L sequence pairs between 14 species of Cucurbitaceae family

	Species*	1	2	3	4	5	6	7	8	9	10	11	12	13
1	*C. pepo*													
2	*C. maxima*	0.00												
3	*C. argyrosperma*	0.00	0.00											
4	*C. moschata*	0.00	0.00	0.00										
5	*L. siceraria*	0.01	0.00	0.00	0.00									
6	*C. lanatus*	0.01	0.00	0.00	0.00	0.00								
7	*C. sativus*	0.00	0.01	0.01	0.01	0.01	0.01							
8	*C. melo*	0.02	0.01	0.01	0.01	0.01	0.01	0.01						
9	*M. charantia*	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.24					
10	*M. cochinchinensis*	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.23				
11	*L. aegyptiaca*	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.23	0.03			
12	*L. acutangula*	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.23	0.03	0.01		
13	*B. hispida*	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.23	0.03	0.02	0.01	
14	*S. edule*	0.03	0.03	0.03	0.03	0.02	0.02	0.03	0.03	0.23	0.03	0.02	0.02	0.02

Table 4. Estimates of evolutionary divergence in *mat*K sequence pairs between 15 species of Cucurbitaceae family

	Species*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*C. pepo*														
2	*C. maxima*	0.00													
3	*C. argyrosperma*	0.00	0.00												
4	*C. moschata*	0.00	0.00	0.00											
5	*L. siceraria*	0.01	0.01	0.01	0.01										
6	*C. lanatus*	0.02	0.02	0.02	0.02	0.01									
7	*C. sativus*	0.02	0.02	0.02	0.02	0.02	0.02								
8	*C. melo*	0.02	0.02	0.02	0.02	0.02	0.02	0.02							
9	*M. charantia*	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02						
10	*M. balsamina*	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01					
11	*M. cochinchinensis*	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02	0.02				
12	*L. aegyptiaca*	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02			
13	*L. acutangula*	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.00		
14	*B. hispida*	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02	
15	*S. edule*	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.03

Substitution bias consisting of transition and transversion at codon position for each luster could reveal the trend of evolution. In this study, the substitution of different bases in analyzed regions is evaluated on entire codon positions (1st+ 2nd + 3rd nucleotide) and shown in Table 5. In general, transitional substitution is higher than transversional substitution in both *rbc*L and *mat*K regions. *rbc*L region shown a higher substitution rate from T to C and A to G in comparison to those of *mat*K region. In contrast, the changing frequency from G to A and C to T of *mat*K is higher than that of *rbc*L.

Estimation of species resolution

The resolution capacity of a barcode is its ability to differentiate and identify species that rely on interspecific differences among DNA sequences. The species is considered as resolved if its individuals construct a specific monophyletic brand (Sikdar et al. 2018). Based on this criterion, in *rbc*L region, five of 14 species consisting of *S. edule*, *M. cochinchinensis*, *L. aegyptiaca*, *C. melo*, and *C. pepo* are resolved and for clarity, bootstrap value lower than 50% is hidden from the figure (Figure 1). Similarly, in *mat*K region, five of 15 species are resolved to consist of *M. balsamina*, *M. cochinchinensis*, *M. charantia*, *S. edule* and *C. sativus* (Figure 2). Thus, both *rbc*L and *mat*K are suitable for identifying *S. edule* and *M. cochinchinensis*, other species are just fit for either *rbc*L or *mat*K. The low species resolution could be due to the small sample size. In order to increase the resolving ability of these two regions, further studies with higher sequence numbers are required (Sikdar et al. 2018). In 2016, a study on the ability of *rbc*L on phylogenetic study of 16 species of *Setaria* genus, the result shows that the *rbc*L gene is highly conserved at the intraspecies level and is not able to differentiate species of *Setaria* genus (Singh et al. 2016). On the contrary, another investigation on 17 cultivars and species of *Prunus* genus presented that *rbc*L is a good marker to study relationships within and among different species of *Prunus* genus (Sarhan et al. 2016).

Table 5. Pattern of nucleotide substitution of *rbc*L and *mat*K regions (in percentage)

	***rbc*L**				***mat*K**			
	A	**T**	**C**	**G**	**A**	**T**	**C**	**G**
A	-	*6.44*	3.88	**12.48**	-	*7.19*	*3.17*	**8.65**
T	*5.87*	-	**11.28**	*4.6*	*6.15*	-	**10.35**	*2.82*
C	*5.87*	**18.71**	-	*4.6*	*6.15*	**23.47**	-	*2.82*
G	**15.94**	*6.44*	*3.88*	-	**18.86**	*7.19*	*3.17*	-

Note: Each entry shows the probability of substitution from one base (row) to another base (column). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversions substitutions are shown in italics)

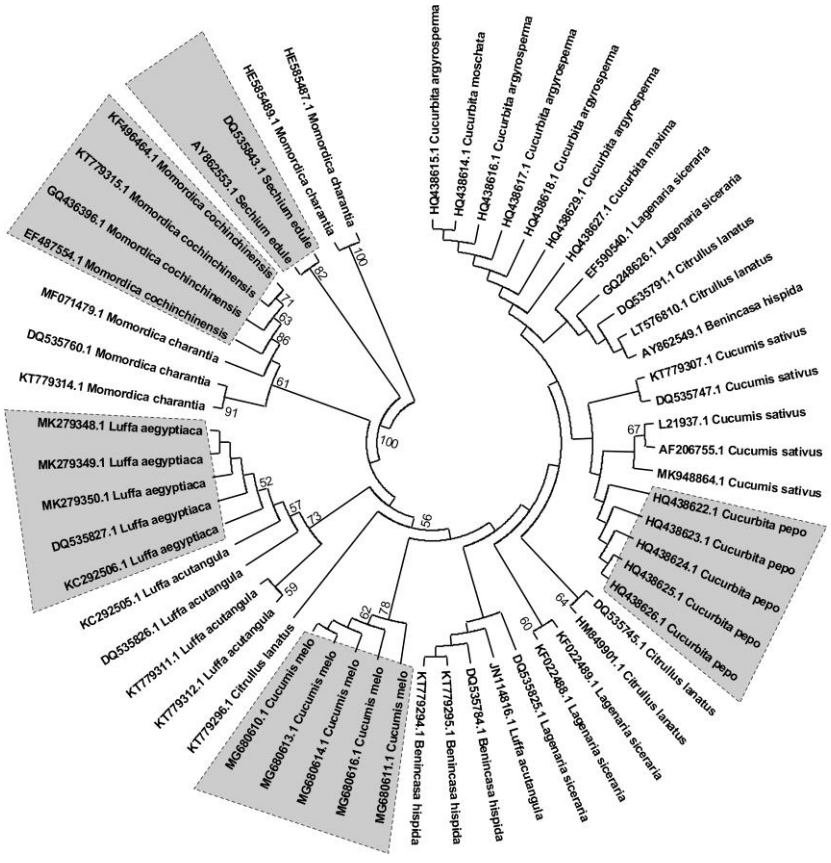

Figure 1. Neighbor-joining tree with 1000 bootstrap replicated based on *rbc*L sequences (Shaded areas present the completely resolved species)

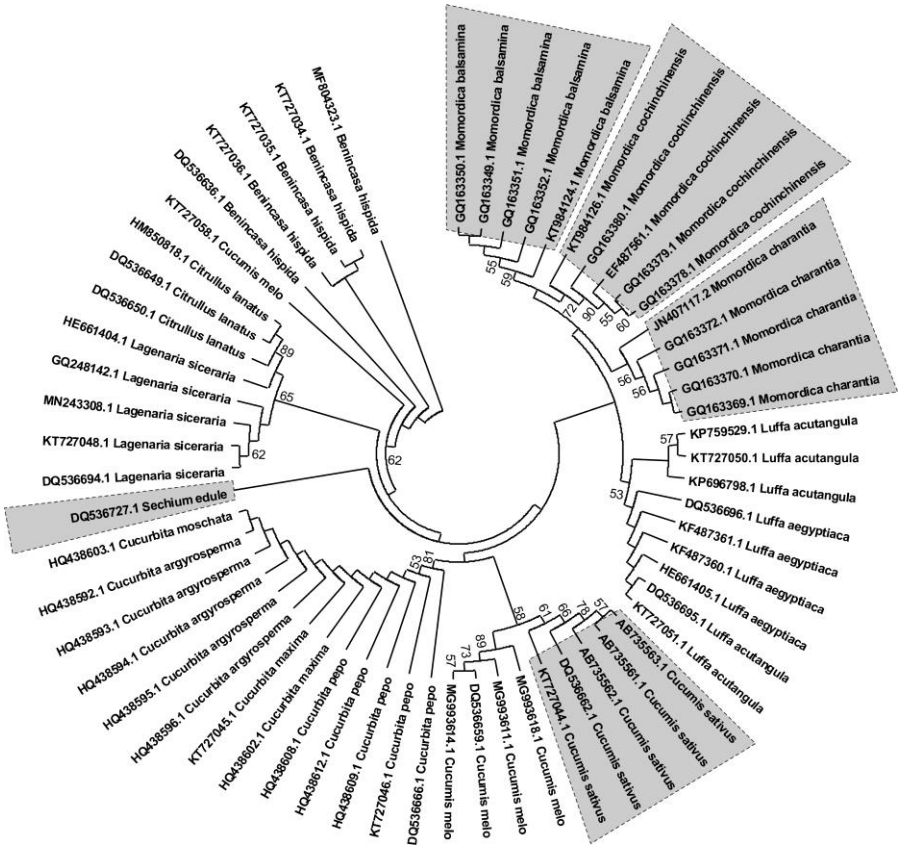

Figure 2. Neighbor-joining tree with 1000 bootstrap replicated based on *mat*K sequences (Shaded areas present the completely resolved species

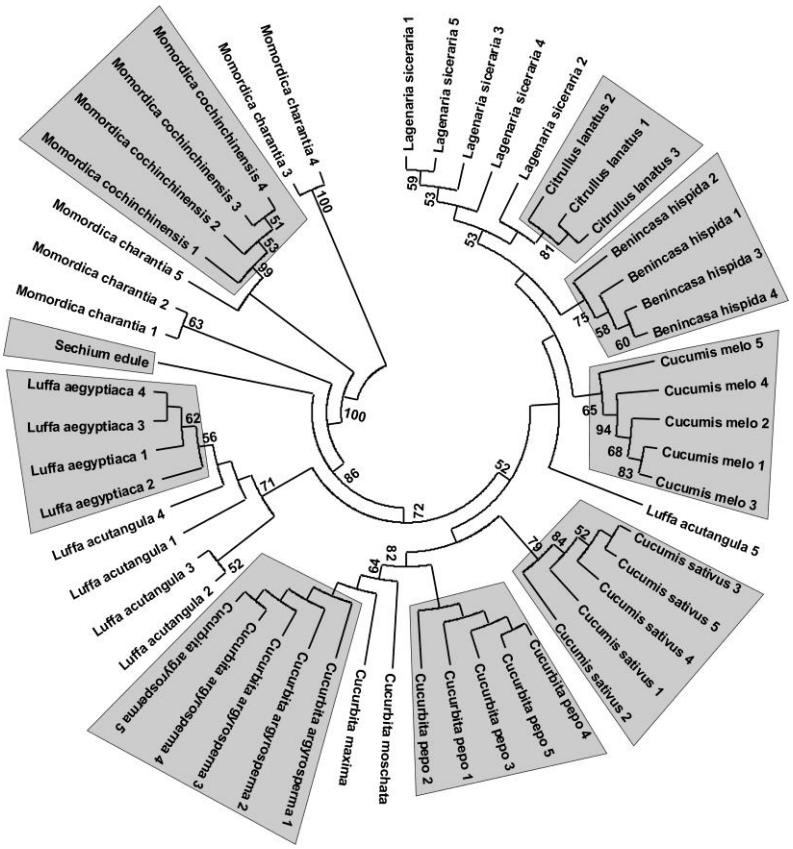

Figure 3. Neighbor-joining tree with 1000 bootstrap replicated based on *rcbL+ mat*K sequences (Shaded areas present the completely resolved species

Since neither *rbc*L nor *mat*K can discriminate all species in this study, further testing was conducted by combining one *rbc*L and one *mat*K sequence of each species into one sequence and targeted for phylogenetic analysis then the result is shown in Figure 3. It can be seen that the species resolutions are clearer than single barcode region. When combining *rbc*L and *mat*K sequences, up to 9 species are resolves namely *C. lanatus, B. hispida, C. melo, C. sativus, C. pepo, C. agryrosperma, L. aegyptiaca, S. edule*, and *M. cochinchinensis.* The high capacity of combining multiple DNA barcode regions in identifying plants have been reported by several studies such as on *Paphiopedilum* (Vu et al. 2019), Araucaria and Inga (Hollingsworth et al. 2009), temperate flora (Burgess et al. 2011). However, the species discrimination efficiency of this combination might be low in some cases such as tropical tree species (Gonzalez et al. 2009). A research group from China has also reported different DNA barcode combinations to show higher species resolution in study of *Gentiana* genus (Liu et al. 2016). Single utilization of *rbc*L, *mat*K, or *rbc*L + *mat*K combination is found not effective for discriminating species in *Calligonum* genus (Li et al. 2014).

The obtained results suggest that species resolving ability of *rbc*L and *mat*K loci are variable among different species of Cucurbitaceae family and the combination of *rbc*L and *mat*K loci could significantly increase resolving power. The limited availability of sequences from some species could affect the precision of the study. Thus, further in silico study should be performed with a higher sequence number to raise the identification ability of DNA barcode loci which could increase the usefulness in conservation and breeding programs of species in Cucurbitaceae family.

REFERENCES

Ahmedand THM, Mohamed ZMA. 2014. Genetic diversity of mango (*Mangifera indica* L.) cultivars in Shendi Area. Ext J Appl Sci 3 (6): 219-224.

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SCH. 2011. Discriminating plant species in a local temperate flora using the *rbc*L + *mat*K DNA barcode. Methods Ecol Evol 2: 333-340.

CBOL Plant Working Group et al. 2009. A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794-12797.

Fatima T, Srivastava A, Somashekar PV, Hanur VS, Rao MS. 2019. Development of DNA-based species identification and barcoding of three important timbers. Bull Natl Res Cent 43: 76. DOI: 10.1186/s42269-019-0116-8.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Petronelli P, Riera B, Roger A, Thhebaud C, Chave J. 2009. Identification of Amazonian trees with DNA barcodes. PLoS ONE 4: e7483. DOI: 10.1371/journal.pone.0007483.

Hebert PDN, Ratnasingham S, deWaard JR 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergence among closely related species. Proc Biol Sci 270 (1): S96-9. DOI: 10.1098/rsbl.2003.0025.

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollinsworth P. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9: 439-457. DOI: 10.1111/j.1755-0998.2008.02439.x

Huxley-Jones EL, Shaw JL, Fletcher C, Parnell J, Watts PC 2012. Use of DNA barcoding to reveal species composition of convenience seafood. Conserv Biol 26 (2): 367-71. DOI: 10.1111/j.1523-1739.2011.01813.x

Kates HR, Soltis PS, Soltis DE. 2017. Evolutionary and domestication history of *Cucurbita* (pumpkin and squash) species inferred from 44 nuclear loci. Mol Phylogenet Evol 111: 98-109.

Kiewnick S, Holterman M, van den Elsen S, van Megen H, Frey JE, Helder J. 2014. Comparison of two short DNA barcoding loci (COI and COII) and two longer ribosomal DNA genes (SSU; LSU rRNA) for specimen identification among quarantine root-knot nematodes (*Meloidogyne* spp.) and their close relatives. Eur J Plant Pathol 140 (1): 97-110.

Knebelsberger T, Landi M, Neumann H, Kloppmann M, Sell AF, Campbell PD, et al. 2014. A reliable DNA barcode reference library for the identification of the North European shelf fish fauna. Mol Ecol Resour 14 (5): 1060-1071. DOI: 10.1111/1755-0998.12238.

Kocyan A, Zhang LB, Schaefer H, Rener SS 2007. A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. Mol Phylogenet Evol 44: 553-577.

Kumar SJU, Rmakrish M, Seethapathy GS, Krishna V, Shaanker R, Ravikanth G. 2020. DNA barcoding of *Momordica* species and assessment of adulteration in Momordica herbal products, an anti-diabetic drug. Plant Gene 22: 100227 DOI: 10.1016/j.plgene.2020.100227.

Li Q, Wu J, Wang Y, Lian X, Wu F, Zhou L, Huan Z, Zhu S. 2017. The phylogenetic analysis of *Dalbergia* (Fabaceae: Papilionaceae) based on different DNA barcodes. Holzforschung 71 (12): 939-949. DOI 10.1515/hf-2017-0052.

Li Y, Feng Y, Wang XY, Liuu B, Lv GH. 2014. Failure of DNA barcoding in discriminating *Calligonum* species. Nord J Bot 32: 511-517.

Lijtmaer DA, Kerr KC, Stoeckle MY, Tubaro PL. 2012. DNA barcoding birds: from field collection to data analysis. Methods Mol Biol. 858: 127-52. DOI: 10.1007/978-1-61779-591-67.

Liu J, Yan HF, Ge XJ. 2016. The use of DNA barcoding on recently diverged species in the Genus *Gentiana* (Gentianaceae) in China. PLoS One 11 (4): e0153008. DOI: 10.1371/journal.pone.0153008.

Mustafa KM, Ewadh MJ, Al-Shuhaib MBS, Hasan HG. 2018. The in silico prediction of the chloroplast maturase K gene polymorphism in several barley varieties. Agriculture (Pol'nohospodarstvo) 64 (1): 3-16.

Saarela JM, Sokoloff PC, Gillespie LJ, Consaul LL, Bull RD. 2013. DNA barcoding the Canadian Arctic Flora: Core plastid barcodes (*rbc*L + *mat*K) for 490 Vascular Plant Species. PLoS ONE 8 (10): e77982. DOI: 10.1371/journal.pone.0077982.

Sarhan S, Hamed F, Al-Youssef A. 2016. The *rbc*L gene sequence variation among and within *Prunus* species. J Agric Sci Tech 18: 1105-1115.

Sikdar S, Tiwari S, Thakur VV, Sapre S, 2018. An in silico approach for evaluation of *rbc*L and *mat*K loci for DNA barcoding of Fabaceae family. Intl J Chem Stud 6 (6): 2446-2451.

Singh A, Drishti, Negi G, Mohanty A. 2016. In silico analysis of sequence variation in *rbc*L gene to assess phylogenetic relations in Setaria species. International Conference on Innovative Research in Agriculture, Food Science, Forestry, Horticulture, Aquaculture, Animal Sciences, Biodiversity, Ecological Sciences and Climate Change (AFHABEC-2016), 22 October 2016, Jawaharlal Nehru University, New Delhi, India.

Suesatpanit T, Osathanunkul K, Madesis P, Osathanunkul M. 2017. Should DNA sequence be incorporated with other taxonomical data for routine identifying of plant species? BMC Compl Altern Med 17: 437. DOI 10.1186/s12906-017-1937-3.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA 6: Molecular evolutionary genetic analysis version 6. Mol Biol Evol 30 (12): 2725-2729.

Tegally A, Jaufeerally-Fakim Y, Dulloo ME. 2019. Molecular characterisation of *Solanum melongena* L. and the crop wild relatives, *S. violaceum* Ortega and *S. torvum* Sw., using phylogenetic/DNA barcoding markers. Genet Resourc Crop Evol. DOI: 10.1007/s10722-019-00827-0.

Thitikornpong W, Palanuvej C, Ruangrungsi N. 2018. DNA barcoding for authentication of the endangered plants in Genus *Aquilaria*. Thai J Pharm Sci 42 (4): 214-220.

Vu TH, Huynh P, Tran HD, Le L. 2019. In silico study on molecular sequences for identification of *Paphiopedilum* species. Evol Bioinform 14: 1-9.

Welbaum GE 2014. Vegetable Production and Practices. CABI Publisher, UK.