BIODIVERSITAS
Volume 22, Number 10, October 2021
Pages: 4216-4222

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d221012

Identification of dragon fruit (*Selenicereus*) species in Mekong Delta based on DNA barcode sequences

TRAN GIA HUY[1], TRAN THANH MEN[2], NGUYEN PHAM ANH THI[1], DO TAN KHANG[1,♥]

[1]Department of Molecular Biotechnology, Biotechnology Research and Development Institute, Can Tho University. 3/2 Street, Ninh Kieu District, Can Tho City, Viet Nam. Tel./fax. +84-919-813035, ♥email: dtkhang@ctu.edu.vn

[2]Department of Biology, School of Natural Sciences, Can Tho University. 3/2 Street, Ninh Kieu District, Can Tho City, Viet Nam

Manuscript received: 11 August 2021. Revision accepted: 18 September 2021.

Abstract. *Huy TG, Men TT, Thi NPA, Khang DT. 2021. Identification of dragon fruit (*Selenicereus*) species in Mekong Delta based on DNA barcode sequences. Biodiversitas 22: 4216-4222. Selenicereus* species is one of the valuable fruiting plants in Vietnam due to its properties, e.g., rich nutrition and medicine activity. Eight DNA barcodes applied to the discrimination power of dragon fruit species in the Mekong delta. Total DNA extracted from fresh roots and the loci of interest were amplified and sequenced. DNA sequences were aligned and determined variable regions. The findings revealed that four loci, including *mat*K, *rbc*L, *rpo*C1, and *atp*F-H reached high PCR yield and specificity compared to those in *ycf*1b, *psb*K-I, and ITS. The *atp*F-H was the most variable region due to the number of single nucleotide polymorphisms (SNPs) and indel mutations, whereas *rpo*C1 was the least one. Based on sequence characteristics, each locus only discriminated some of the *Selenicereus monacanthus* from Southern Horticultural Research Institute identified by combining three loci, *atp*F-H, *mat*K, and *rbc*L. The results elucidated the close genetic relationship between Mekong delta dragon fruits and National Center for Biotechnology Information (NCBI) database. Furthermore, this finding generated a DNA barcode database of ten dragon fruit accessions and suggested that multiple loci in the chloroplast genome should be a reliable solution for identifying this highly commercial fruiting plant.

Keywords: *atp*F-H, *mat*K, plant authentication, *rbc*L, *Selenicereus*

INTRODUCTION

Selenicereus (A. Berger) Britton & Rose, pitahaya or dragon fruit, belongs to the family Cactaceae, is distributed from South America regions. Individuals of this genus are described as vine cacti with unique traits such as climbing, aerial roots, three angled stems, and glabrous large-scaled berry (Montoya-Arroyo et al. 2014). It is grown as ornamental plants in gardens and indoors for its size, aromatic, and abnormal time blooming flowers. At present, dragon fruits are being exported globally and have a high economic value as exotic fruit crops in harsh regions where water is limited. The Crassulacean acid metabolism pathway is utilized for carbon dioxide fixation and is highly tolerant to water stress (Ibrahim et al. 2018). The fruits have played a remarkable role in medicine, food, and ornamentally. Dragon fruit is rich in essential nutrients such as vitamins, minerals, complex carbohydrates, dietary fibers, and antioxidants (Wichienchot et al. 2010; Tenore et al. 2012). Dragon fruit can withstand prolonged drought. Therefore, it considers as a high potential fruit for horticultural development, especially in areas where drought is a limiting factor for other fruits. There are 14 species of dragon fruit in the world (Cisneros and Tel-Zur 2012). White flesh dragon fruit (*Selenicereus undatus*) and red flesh dragon fruit (*Selenicereus monacanthus*) cultivate in Vietnam. To the best of our knowledge, dragon fruits were classified into the *Selenicereus* genus base on DNA sequences in both nuclear and plastid (Korotkova et al. 2017). Thus, the accepted scientific name of dragon fruit is *Selenicereus* instead of *Hylocereus*.

DNA barcodes are short DNA fragments, around 400-800 bp, found in all plant species (DeSalle and Goldstein 2019). DNA is a specialized chain of "letters" that distinguish between organisms and/or individuals despite very similar morphological features between them. By integrating the advances of molecular biology, sequencing technologies, and bioinformatics, DNA barcodes provide a quick and accurate means to recognize previously known, described, and classified species and construct a DNA database for them (Kress 2017). As a result of the low nucleotide substitution rate in land plants, it is a challenge to determine a standard DNA barcode (Fazekas et al. 2012). Furthermore, several loci from the plastid genome analyze for species discrimination power. Two genes of *mat*K and *rbc*L and their combination proposed as a standard DNA barcode for land plants (DeSalle and Goldstein 2019).

In *Selenicereus* species, some studies focused on the morphology and Inter-simple sequence repeat (ISSR marker) for genetic diversity and species identification (Tao et al. 2014; Abirami et al. 2021). However, morphological traits influenced by environmental factors and the band pattern may not be reliable because ISSR is a dominant marker. Based on nucleotide sequences, which is environment independence, DNA barcodes proposed a potential solution for plant authentication. This current study was to disclose the ability of DNA barcodes to identify *Selenicereus* species in The Mekong delta.

MATERIALS AND METHODS

Plant sampling

Ten samples of *Selenicereus* species were collected from Southern Horticultural Research Institute (SOFRI) and fruit gardens in The Mekong delta (Table 1). The *S. monacanthus* DF1 was the high yield variety cultivated ubiquitously in The Mekong delta. Such samples were verified by SOFRI staff.

DNA extraction and amplification

Fresh roots sterilized with 70% ethanol and then cut into tiny pieces. Samples were incubated at liquid nitrogen for 5 minutes and grind into powder. Total DNA was extracted using the CTAB-based protocol (Roger and Bendich 1988) with appropriate modifications. The quantity and purity of DNA were measured by Nanodrop spectrophotometer 2000C (Thermo Scientific, USA). DNA integrity was examined by 1% agarose electrophoresis.

Eight loci, including four genes *mat*K, *rpo*C1, *rbc*L, *ycf*1b, and four noncoding spacers *psb*A-trnH, *atp*F-H, *psb*K-I, ITS were amplified by 30 µL of volume reaction. The reagents consist of 15 µL of master mix 2X (Bioline, United Kingdom), 0.4 µM of each forward and reverse primer, and DNA template. Primer sequences are listed in Table 2. The amplification process was performed in C1000 thermocycler (Bio-rad, USA) (Table 2). PCR products were separated by 2% agarose electrophoresis for 45 minutes at 50V, and the bands were visualized by Run-Safe stain (Cleaver Scientific, United Kingdom). Amplicons with clear bands and no nonspecific products were submitted to Nextgen Biotechnology corporation for sequencing (3500 Genetic Analyzer, Applied Biosystem).

Table 1. Main characteristics of ten dragon fruit samples in this study

Code	Species	Shape	Source	Code	Species	Shape	Source
DF1	*S. monacanthus*	Red flesh, oval fruit	SOFRI, cultivated	DF6	*S. undatus*	White flesh, oval fruit	Ben Tre, cultivated
DF2	*S. monacanthus*	Red flesh, oval fruit	Ben Tre, cultivated	DF7	*S. undatus*	White flesh, oval fruit	Ca Mau, cultivated
DF3	*S. monacanthus*	Red flesh, oval fruit	Ca Mau, cultivated	DF8	*S. megalanthus*	Yellow skin, white flesh, oval fruit	Dong Thap, cultivated
DF4	*Selenicereus* sp.	Red flesh, round fruit	Ca Mau, cultivated	DF9	*S. megalanthus*	Yellow skin, white flesh, oval fruit	Tien Giang, cultivated
DF5	*Selenicereus* sp.	Purple flesh, oval fruit	SOFRI, cultivated	DF10	*Selenicereus* sp.	Not available	An Giang, wild

Note: DF1: *S. monacanthus* SORFI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SORFI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: *Selenicereus* sp. AG.

Table 2. Nucleotide sequences of primer pairs for amplification of DNA barcode candidates (Primer database from boldsystems.org)

Locus	Sequence (5'-3')	Thermal cycle (35 cycles)
*mat*K	F: CGATCTATTCATTCAATATTTC R: TCTAGCACACGAAAGTCGAAGT	94°C-1 min; 94˚C- 30 sec, 50°C-40 sec, 72°C-40 sec; 72°C-5 min
*rbc*L	F: ATGTCACCACAAACAGAGACTAAAGC R: GTAAAATCAAGTCCACCRCG	94°C-4 min; 94°C- 30 sec, 55°C-30 sec, 72°C-1 min; 72°C-10 min
*psb*A-*trn*H	F: GTTATGCATGAACGTAATGCTC R: CGCGCATGGTGGATTCACAATCC	
*rpo*C1	F: GGCAAAGAGGGAAGATTTCG R: CCATAAGCATATCTTGAGTTGG	94°C-4 min; 94°C- 30 sec, 58°C-30 sec, 72°C-1 min; 72°C-10 min
*atp*F-*atp*H	F: ACTCGCACACACTCCCTTTCC R: GCTTTTATGGAAGCTTTAACAAT	94°C-4 min; 94°C- 30 sec, 51°C-40 sec, 72°C-40 sec; 72°C-5 min
*psb*K-*psb*I	F: TTAGCCTTTGTTTGGCAAG R: TTAGCCTTTGTTTGGCAAG	
ITS	F: TCCGGGAACCTGCGG R: TCCTCCGCTTTGATGC	95°C-5 min; 95°C- 30 sec, 57°C-30 sec, 72°C-1 min; 72°C-5 min
*ycf*1b	F: TCTCGACGAAAATCAGATTGTTGTGAAT R: ATACATGTCAAGTGATGGAAAA	95°C-5 min; 95°C- 30 sec, 51°C-40 sec, 72°C-1 min; 72°C-10 min

Table 3. Accession numbers for DNA barcode sequences in this study

Samples	*atp*F-H	*rbc*L	*rpo*C1	*mat*K
DF1	OK094559	OK094539	OK094549	OK094529
DF2	OK094560	OK094540	OK094550	OK094530
DF3	OK094561	OK094541	OK094551	OK094531
DF4	OK094562	OK094542	OK094552	OK094532
DF5	OK094563	OK094543	OK094553	OK094533
DF6	OK094564	OK094544	OK094554	OK094534
DF7	OK094565	OK094545	OK094555	OK094535
DF8	OK094566	OK094546	OK094556	OK094536
DF9	OK094567	OK094547	OK094557	OK094537
DF10	OK094568	OK094548	OK094558	OK094538

Note: DF1: *S. monacanthus* SORFI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SORFI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: *Selenicereus* sp. AG

Data analysis

Raw sequences were interpreted by Bioedit software version 7.2.1. The quality value checks to confirm the accuracy of the sequencing procedure, and unidentified nucleotide was verified based on the chromatogram. These sequences were submitted to Genbank for accession number registration (Table 3). Multiple sequence alignment was conducted by following the ClustalW algorithm. Conservative and variable regions determine by MEGA X software (Kumar et al. 2018). Homologous sequences belong to the *Selenicereus* genus on Nation Center Biotechnology Information (NCBI) were collected by Basic Local Alignment Search Tool (BLAST).

RESULTS AND DISCUSSION

Amplification of DNA barcode sequences

Eight DNA barcode sequences, including four protein-encoding genes (*ycf*1b, *rbc*L, *rpo*C1, and *mat*K) and four non-coding sequences (*atp*F-H, *psb*A-*trn*H, *psb*K-I, and ITS), were amplified and visualized on the gel (Figure 1). From the gel patterns, amplicon size ranged from 400 to 900 bp for eight loci. It indicated that three genes *rpo*C1, *rbc*L, *mat*K, and two spacers *atp*F-H, *psb*A-*trn*H showed unique bands, reflecting the PCR specificity. No band has appeared in the negative control sample, indicating that external contamination under-controlled. Therefore, such five loci were suitable to sequence and analyze species identification. On the other hand, the presence of nonspecific bands from the *ycf*1b gene, ITS, and *psb*K-I spacer made such loci unsuitable for sequencing.

Sequence analysis for DNA barcode candidates

Based on the chromatogram, the quality value of DNA sequences was verified. The *psb*A-*trn*H spacer indicated multiple unidentified characters while *atp*F-H, *mat*K, *rpo*C1, and *rbc*L expressed clear nucleotide signals. Conservative and variable sites of such loci were detected based on multiple sequence alignment (Table 4). Based on the alignment result, all sequences showed a high amount of conserved sites. As a result, *Selenicereus* species expressed a close genetic relationship. However, the occurrence of SNPs and indel-mutations revealed the nucleotide diversity of such dragon fruit species. The *atp*F-H spacer was the highest variable locus with six variable sites. By contrast, three genes consist of *mat*K, *rpo*C1, and *rbc*L were highly similar. In terms of indel-mutations, nine mutations were detected from the *atp*F-H sequence, while the frequency of this mutation was quite low in three other loci. Thus, it could suggest that the noncoding sequence was more variable than coding sequences and proposed as a potential barcode sequence for species discrimination.

A serial insertion of 9 nucleotides ATTAGGTAC was found on *Selenicereus* sp. DF4 (red flesh and round fruit) (Table 5). Because of this mutation, DF4 distinguish from other red flesh dragon fruits. This result confirmed that *atp*F-H spacer was highly polymorphic and able to identify this dragon fruit variety. Santos and Pereira (2018) analyzed more than 44,000 sequences belonging to 206 different plant families. Their findings suggested that the *atp*F-H spacer was a suitable region for SPInDel (Species Identification by Insertions/Deletions) concept. Furthermore, five substitution mutations were detected and classified into three SNPs as T/C, C/A, and G/A.

Sequence analysis from the *rbc*L gene reflected that this plastid gene was valuable for species delimitation (Table 6). The *S. monacanthus* DF1 (red flesh) and *Selenicereus* sp. DF5 (purple flesh) from SOFRI were identified successfully from the other samples by four valuable SNPs, including G/C, A/C, G/T, and T/C, and two consecutive insertions with A and G. The *rbc*L gene was also evaluated as a hypervariable region, which is the promising DNA barcode to identify *S. monacanthus* and three species *S. annthonyanus*, *S. grandifloras* and *S. validus* (Qin et al. 2021).

The *rpo*C1 gene, a coding sequence in the plastid genome, revealed the incorrect identification of dragon fruits (Table 7). For instance, samples DF1, DF2, and DF3 belonged to *S. monacanthus*, but their sequence was variable. Furthermore, four individuals DF1, DF5, DF6, and DF10 were different species, and their nucleotide sequences were identical. Although *rpo*C1 was one of 10 informative markers for phylogeny in Cactaceae (Köhler et al. 2020), this plastid gene was not a suitable barcode for *Selenicereus* species.

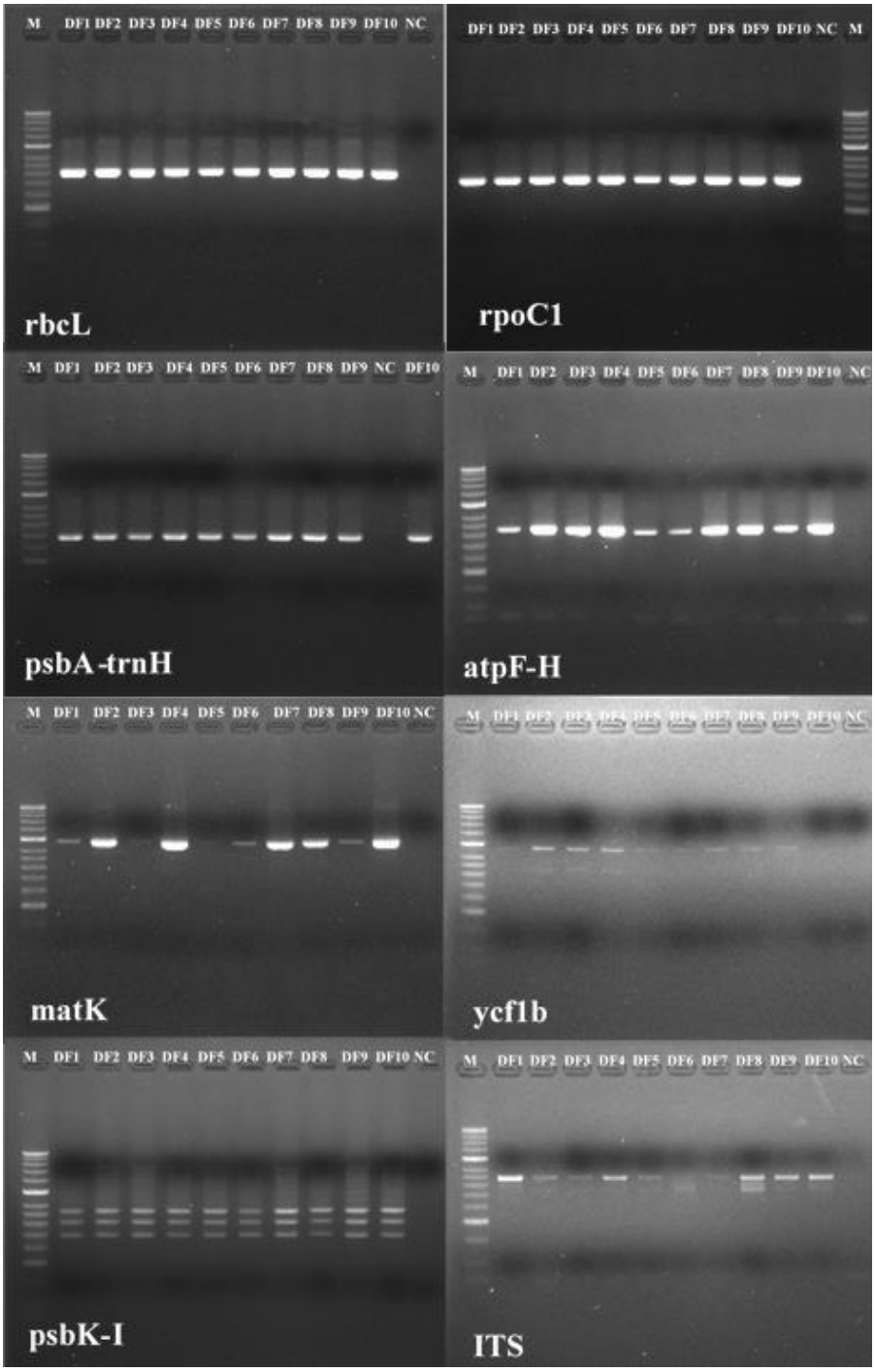

Figure 1. Amplicons of eight DNA barcode locus on 2% agarose gel. Note: M: 50 bp hyperladder (Bioline, England); DF1: *S. monacanthus* SOFRI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SOFRI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: Selenicereus sp. AG; NC: Negative control.

Table 4. Sequence characteristics of four DNA barcode candidates

Locus	Aligned length (bp)	Conserved sites	Variable sites		Indel mutations
			Singleton sites	Parsimony informative sites	
*atp*F-H	585	571	5 (0.85%)	1 (0.17%)	9 (1.54%)
*mat*K	782	778	1 (0.13%)	3 (0.38%)	0 (0%)
*rbc*L	506	502	1 (0.20%)	3 (0.59%)	2 (0.40%)
*rpo*C1	452	450	1 (0.22%)	1 (0.22%)	2 (0.44%)

Table 5. Single nucleotide polymorphisms (SNPs) and indel mutations of *atp*F-H spacer

Accession number	Samples	Position														
		18	18	18	18	27	27	28	28	28	28	28	28	49	59	
		1	2	3	4	8	9	0	1	2	3	4	5	6	5	7
OK094559	DF1	**T**	**T**	**T**	**G**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**C**	**T**
OK094560	DF2	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**
OK094561	DF3	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**
OK094562	DF4	**C**	**A**	**A**	**A**	**A**	**T**	**T**	**A**	**G**	**G**	**T**	**A**	**C**	**A**	**C**
OK094563	DF5	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**
OK094564	DF6	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**
OK094565	DF7	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**C**
OK094566	DF8	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**
OK094567	DF9	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**
OK094568	DF10	**.**	**.**	**.**	**.**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**_**	**.**	**.**

Note: DF1: *S. monacanthus* SORFI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SORFI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: *Selenicereus* sp. AG.

Table 6. Single nucleotide polymorphisms (SNPs) and indel mutations of *rbc*L gene

Accession number	Sample	Positions					
		194	241	489	490	491	492
OK094539	DF1	**G**	**A**	**A**	**G**	**G**	**T**
OK094540	DF2	**.**	**C**	**_**	**_**	**.**	**.**
OK094541	DF3	**.**	**C**	**_**	**_**	**.**	**.**
OK094542	DF4	**.**	**C**	**_**	**_**	**.**	**.**
OK094543	DF5	**.**	**.**	**.**	**.**	**.**	**.**
OK094544	DF6	**C**	**.**	**_**	**.**	**T**	**C**
OK094545	DF7	**.**	**C**	**_**	**_**	**.**	**.**
OK094546	DF8	**.**	**.**	**_**	**_**	**.**	**.**
OK094547	DF9	**.**	**.**	**_**	**_**	**.**	**.**
OK094548	DF10	**.**	**C**	**_**	**_**	**.**	**.**

Note: DF1: *S. monacanthus* SORFI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SORFI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: *Selenicereus* sp. AG.

Table 7. Single nucleotide polymorphisms (SNPs) and indel mutations of *rpo*C1 gene

Accession number	Sample	Positions			
		2	4	6	18
OK094549	DF1	**A**	**T**	**A**	**A**
OK094550	DF2	**C**	**.**	**_**	**_**
OK094551	DF3	**.**	**.**	**.**	**_**
OK094552	DF4	**.**	**.**	**.**	**_**
OK094553	DF5	**.**	**.**	**.**	**.**
OK094554	DF6	**.**	**.**	**.**	**.**
OK094555	DF7	**C**	**G**	**.**	**_**
OK094556	DF8	**C**	**.**	**_**	**_**
OK094557	DF9	**.**	**.**	**.**	**.**
OK094558	DF10	**.**	**.**	**.**	**.**

Note: DF1: *S. monacanthus* SORFI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SORFI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: *Selenicereus* sp. AG.

It illustrated that the *mat*K sequence showed nucleotide polymorphism; however, this gene was non-informative to identify dragon fruits species (Table 8). Data from six dragon fruits species from Northern Vietnam (Huong et al. 2021) also showed that the *mat*K gene contained variable sites, but these nucleotides could not discriminate against a particular species. On the other hand, the deletion of adenine in *Selenicereus* sp. DF5 compared to *S. monacanthus* DF1 SOFRI (position 817) made these two varieties distinguishable.

In comparison with dragon fruit sequences from the NCBI database, it showed that five species, *S. undatus* (NC_053698.1), *S. tricae* (LT745724.1), *S. ocamponis* (LT745687.1) (Korotkova et al. 2017), *S. costaricensis* (JQ590992.1) and *S. peruvianus* (AY015310.1) showed the identity percent ranging from 97-100% compared with the *S. monacanthus* SOFRI (Table 9). *Selenicereus costaricensis* is the purple flesh dragon fruit variety in Costa Rica (Viñas and Jiménez 2016), this species contained a high amount of bioactive compounds with no cytotoxic effects (Paśko et al. 2021). *Selenicereus tricae* are the wild dragon fruit variety in Belize (Gómez-Hinostrosa et al. 2014) and *S. ocamponis* is the red flesh dragon fruit variety in Mexico (Ibrahim et al. 2018). The high identity percentage indicated a close genetic relationship among these accessions with *S. monacanthus* SOFRI. Based on sequence similarity, three loci, including *mat*K, *atp*F-H, and *rbc*L could discriminate *S. monacanthus* SOFRI from dragon fruits in other countries.

The maturase *mat*K is one of the fastest evolving genes in the chloroplast genome. It was widely applied for phylogenetic studies at the species level since it played a crucial role in the RNA splicing process (Schmitz-Linneweber et al. 2015). The utility of *mat*K region tested for species-level identification on 528 species of Cactaceae, including approximately 75% of Mexican species. Yesson et al. (2011) find that the DNA barcode by *mat*K could identify exactly 77% of collected species. However, the expectation value is more than that percentage, and the change of nucleotides in primer regions is the main

drawback for PCR specificity of such gene (Yesson et al. 2011). Bell et al. (2017) generated the *rbc*L database for the identification of plant mixture. This study succeeded in accurate species-level identification for eight angiosperm species: *Populus tremuloides* (Salicaceae), *Populus deltoides*, *Broussonetia papyrifera* (Moraceae), *Carya illinoinensis* (Juglandaceae), *Bassia scoparia* (Amaranthaceae), *Ambrosia artemisiifolia* (Asteraceae), *Artemisia tridentata* (Asteraceae), *Poa pratensis* (Poaceae) and a family-level identification of *Zea mays* (Poaceae). Thus, the *rbc*L gene should be an improvement on ITS, a nuclear sequence has been utilized widely for plant identification. The *rbc*L (1, 5-ribulose bisphosphate carboxylase/oxygenase large subunit) gene encodes a large subunit of rubisco protein, which is an important enzyme for photosynthesis (Fangru et al. 2020). Thus, *mat*K and *rbc*L were responsible for key metabolism processes in cells, so the changes of their sequences might be powerful markers for species identification.

Frequent plant hybridization via mechanisms such as polyploidy and various breeding systems have given rise to new hybrids formation (Fazekas et al. 2012). Several breeding programs were submitted to increase the yield and quality of dragon fruits lead to high intra- and interspecific hybridization. This phenomenon generated taxonomical confusion, leading to the complexity of dragon fruit species and varieties (Jian-ye et al. 2021). Therefore, a single locus possesses insufficient discrimination power to identify these species. In this study, the *S. monacanthus* SOFRI variety was authenticated by three loci, *atp*F-H + *rbc*L + *mat*K. Several studies contributed that species resolution was improved significantly by multiple loci. Data from 12 plant genera showed that chloroplast genome sequences are highly variable, and such regions should be the priority when seeking the suitable loci to resolve closely related plant species or varieties and for DNA barcoding (Dong et al. 2012). The utilization of multiple loci considers as a potential solution for accurate identification (Fazekas et al. 2012; Wu et al. 2017). A single DNA barcode only identifies some of the 18 species in the Melilotus genus, a herbal plant in North Africa. According to these findings, the combination of five loci, *mat*K + *rbc*L + *trn*L-F + *trn*H-*psb*A + ITS showed the greatest species resolution while the single *rbc*L was the least. Khan et al. (2017) reported that four coding regions *mat*K, *rbc*L, *rpo*B, and *rpo*C1 were highly conservative among the taxa. On the other hand, the sequences of two intergenic spacers *psb*K-*psb*I and *atp*F-*atp*H were variable, specifically identifying the medicinal plant *Rhazya stricta* (Khan et al. 2017). Gogoi and Bhau (2018) indicated that single-locus ITS or combined with plastid *mat*K expressed the better species authentication of the genus Nepenthes based on barcoding gaps. Vu et al. (2020) also suggested that the combination of ITS + *mat*K was the most potential DNA barcode for Vietnamese *Paphiopedilum* species. Therefore, It was reasonable to conclude that reliable identification was strongly supported by more than one DNA barcode locus. Last but not least, an effective locus for this species may have some limitations for another species.

Table 8. Single nucleotide polymorphisms (SNPs) and indel mutations of *mat*K gene

Accession number	Sample	Positions								
		738	784	787	799	800	801	809	817	826
OK094529	DF1	**A**	**A**	**A**	**_**	**A**	**T**	**_**	**A**	**A**
OK094530	DF2	**.**	**.**	**.**	**G**	**.**	**.**	**_**	**.**	**.**
OK094531	DF3	**.**	**C**	**.**	**A**	**T**	**G**	**G**	**.**	**.**
OK094532	DF4	**.**	**.**	**.**	**_**	**.**	**.**	**_**	**.**	**.**
OK094533	DF5	**.**	**.**	**.**	**_**	**.**	**.**	**_**	**_**	**.**
OK094534	DF6	**.**	**.**	**.**	**_**	**.**	**.**	**_**	**_**	**_**
OK094535	DF7	**.**	**.**	**.**	**_**	**.**	**.**	**A**	**.**	**_**
OK094536	DF8	**.**	**.**	**.**	**_**	**.**	**.**	**_**	**.**	**.**
OK094537	DF9	**_**	**.**	**_**	**_**	**.**	**.**	**_**	**.**	**.**
OK094538	DF10	**.**	**.**	**.**	**_**	**.**	**.**	**_**	**.**	**.**

Note: DF1: *S. monacanthus* SORFI; DF2: *S. monacanthus* BT, DF3: *S. monacanthus* CM; DF4: *Selenicereus* sp.; DF5: *Selenicereus* sp. SORFI; DF6: *S. undatus* BT; DF7: *S. undatus* CM; DF8: *S. megalanthus* DT; DF9: *S. megalanthus* TG; DF10: *Selenicereus* sp. AG

Table 9. Nucleotide polymorphism between *Selenicereus monacanthus* SOFRI and other *Selencicereus* species on NCBI database

Locus	Length (nt)	Coverage (%)	Identities (%)	Number of gaps	Accessions
*atp*F-H	576	100	97.61	9	*S. undatus* (NC_053698.1)
*mat*K	782	100	99.74	0	*S. tricae* (LT745724.1)
					S. undatus (NC_053698.1)
					S. peruvianus (AY015310.1)
					S. ocamponis (LT745687.1)
*rbc*L	506	100	99.41	2	*S. costaricensis* (JQ590992.1)
					S. undatus (NC_053698.1)
*rpo*C1	452	100	100	0	*S. undatus* (NC_053698.1)

This study successfully constructed DNA barcodes data for common *Selenicereus* species in the Mekong delta, consisting of four chloroplast loci: *rbc*L, *rpo*C1, *mat*K and *atp*F-H. Such sequences showed high PCR yield and specificity as well as sequence quality. The *rpo*C1 region was not a suitable barcode for this plant because of the highly conservative sequence. Based on the appearance of SNPs and indel mutations, three loci: *rbc*L, *mat*K and *atp*F-H were able to distinguish some species. The integration of these three loci enhances the discrimination power and the successful identification of *S. monacanthus* DF1 varieties.

ACKNOWLEDGEMENTS

Tran Gia Huy, VINIF.2020.ThS.68 was funded by Vingroup Joint Stock Company and supported by the Domestic Master/ PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VINBIGDATA)".

REFERENCES

Abirami K, Swain S, Baskaran V, Venkatesan K, Sakthivel K, Bommayasamy N. 2021. Distinguishing three Dragon fruit (*Hylocereus* spp.) species grown in Andaman and Nicobar Islands of India using morphological, biochemical and molecular traits. Sci Rep 11: 2894. DOI: 10.1038/s41598-021-81682-x

Bell KL, Loeffler VM, Brosi BJ. 2017. An *rbc*L reference library to aid in the identification of plant species mixtures by DNA metabarcoding. Appl Plant Sci 5 (3): 1600110. DOI: 10.3732/apps.1600110

Cisneros A, Tel-Zur N. 2012. Evaluation of interspecific-interploid hybrids (F1) and back crosses (BC1) in *Hylocereus* species (Cactaceae). In: Andrew S (ed) Meiosis: Molecular mechanisms and cytogenetic diversity. InTechOpen, Croatia. DOI: 10.5772/32435

DeSalle R, Goldstein P. 2019. Review and interpretation of trends in DNA barcoding. Front Ecol Evol 7: 302. DOI: 10.3389/fevo.2019.00302

Dong W, Liu J, Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PloS one 7 (4): e35071. DOI: 10.1371/journal.pone.0035071

Fangru N, Yuxin H, Xudong L, Jia F, Junping L, Qi L, Shulian X. 2020. Analysis of adaptive evolution and coevolution of *rbc* L gene in the Genus *Hildenbrandia* (Rhodophyta). Evol Bioinform 16: 1176934320977862. DOI: 10.1177/1176934320977862

Fazekas AJ, Kuzmina ML, Newmaster SG, Hollingsworth PM. 2012. DNA barcoding methods for land plants. In: DNA Barcodes. Humana Press, Totowa. DOI: 10.1007/978-1-61779-591-6_11

Gómez-Hinostrosa C, Hernández HM, Terrazas T, Correa-Cano ME. 2014. Studies on Mexican Cactaceae. V. Taxonomic notes on *Selenicereus tricae*. Brittonia 66 (1): 51-59. DOI: 10.1007/s12228-013-9308-y

Gogoi B, Bhau BS. 2018. DNA barcoding of the genus *Nepenthes* (Pitcher plant): A preliminary assessment towards its identification. BMC Plant Biol 18: 153. DOI: 10.1186/s12870-018-1375-5

Huong NTT, Hung NQ, Lu DV, Khanh TD. 2021. Morphological characteristics and genetic diversity of some red dragon fruit samples in field trials and *mat*K marker sequencing. J Sci Engineer Res 8 (1): 189-195.

Jian-ye C, Fang-fang X, Yan-ze C, Can-bin C, Wang-jin L, Xiao-di H, Yong-hua Q. 2021. A chromosome-scale genome sequence of pitaya (*Hylocereus undatus*) provides novel insights into the genome evolution and regulation of betalain biosynthesis. Hortic Res 8: 164. DOI: 10.1038/s41438-021-00612-0

Köhler M, Reginato M, Souza-Chies TT, Majure LC. 2020. Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. Front Plant Sci 11: 729. DOI: 10.3389/fpls.2020.00729

Korotkova N, Borsch T, Arias S. 2017. A phylogenetic framework for the Hylocereeae (Cactaceae) and implications for the circumscription of the genera. Phytotaxa 327 (1): 1-46. DOI: 10.11646/phytotaxa.327.1.1

Khan SA, Baeshen MN, Ramadan HA, Baeshen NA. 2017. Emergence of plastidial intergenic spacers as suitable DNA barcodes for arid medicinal plant *Rhazya stricta*. Am J Plant Sci 8 (8): 1774-1789. DOI: 10.4236/ajps.2017.88121

Kress WJ. 2017. Plant DNA barcodes: Applications today and in the future. J Syst Evol 55: 291-307. DOI: 10.1111/jse.12254

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35 (6): 1547. DOI: 10.1093/molbev/msy096

Ibrahim SRM, Milborne P, Mohamed GA, Khedr AIM, Zayed MF, El-Kholy AAES. 2018. Genus *Hylocereus*: Beneficial phytochemicals, nutritional importance, and biological relevance—A review. J Food Biochem 42 (2): e12491. DOI: 10.1111/jfbc.12491

Montoya-Arroyo A, Schweiggert RM, Pineda-Castro ML, Sramek M, Kohlus R, Carle R, Esquivel P. 2014. Characterization of cell wall polysaccharides of purple pitaya (*Hylocereus* sp.) pericarp. Food Hydrocolloids 35: 557-564. DOI: 10.1016/j.foodhyd.2013.07.010

Paśko P, Galanty A, Zagrodzki P, Ku YG, Luksirikul P, Weisz M, Gorinstein S. 2021. Bioactivity and cytotoxicity of different species of pitaya fruits – A comparative study with advanced chemometric analysis. Food Biosci 40: 100888. DOI: 10.1016/j.fbio.2021.100888

Qin Q, Li J, Zeng S, Xu Y, Han F, Yu J. 2021. The complete plastomes of red fleshed pitaya (*Selenicereus monacanthus*) and three related *Selenicereus* species: Insights into gene losses, IR expansions and phylogenomic implications. Research Square. DOI: 10.21203/rs.3.rs-373541/v1

Santos C, Pereira F. 2018. Identification of plant species using variable length chloroplast DNA sequences. Forensic Sci Int Genet 36: 1-12. DOI: 10.1016/j.fsigen.2018.05.009

Schmitz-Linneweber C, Lampe MK, Sultan LD, Ostersetzer-Biran O. 2015. Organellar maturases: A window into the evolution of the spliceosome. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1847 (9): 798-808. DOI: 10.1016/j.bbabio.2015.01.009

Tao J, Qiao G, Wen XP, Gao GL, Liu T, Peng ZJ. 2014. Characterization of genetic relationship of dragon fruit accessions (*Hylocereus* spp.) by morphological traits and ISSR markers. Sci Hortic 170: 82-88. DOI: 10.1016/j.scienta.2014.03.006

Tenore GC, Novellino E, Basile A. 2012. Nutraceutical potential and antioxidant benefits of red pitaya (*Hylocereus polyrhizus*) extracts. J Funct Foods 4 (1): 129-136. DOI: 10.1016/j.jff.2011.09.003

Viñas M, Jiménez VM. 2016. Occurrence and characterization of calcium oxalate crystals in stems and fruits of *Hylocereus costaricensis* and *Selenicereus megalanthus* (Cactaceae: Hylocereeae). Micron 89: 21-27. DOI: 10.1016/j.micron.2016.07.002

Vu HT, Vu QL, Nguyen TD, Tran N, Nguyen TC, Luu PN. 2020. Genetic diversity and identification of Vietnamese *Paphiopedilum* species using DNA Sequences. Biology 9 (1): 9. DOI: 10.3390/biology9010009

Wichienchot S, Jatupornpipat M, Rastall RA. 2010. Oligosaccharides of pitaya (dragon fruit) flesh and their prebiotic properties. Food Chem 120 (3): 850-857. DOI: 10.1016/j.foodchem.2009.11.026

Wu F, Ma J, Meng Y, Zhang D. 2017. Potential DNA barcodes for *Melilotus* species based on five single loci and their combinations. PLoS One 12 (9): e0182693. DOI: 10.1371/journal.pone.0182693

Yesson C, Barcenas RT, Hernandez HM, Ruiz-Maqued MDe L, Prado A, Rodriguez VM, Hawkins JA. 2011. DNA barcodes for Mexican Cactaceae, plants under pressure from wild collecting. Mol Ecol Resour 11 (5): 775-783. DOI: 10.1111/j.1755-0998.2011.03009.x