Ind. J. Pharm. Edu. Res., 2023; 57(2s):s335-s344.
https://www.ijper.org

Original Article

Authentication of *Averrhoa carambola* L. Using DNA Barcoding: Detection of SNPs and Evaluation of Phylogenetic Relationship

Alok Kumar Giri, Bhagyashree Biswal, Biswajit Jena, Laxmikanta Acharya*

Molecular Biology and Genetic Engineering Lab, Centre for Biotechnology, School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, INDIA.

ABSTRACT

Background: *Averrhoa carambola* L. (star fruit) is an underutilized plant in the family *Oxalidaceae*. The plant has various pharmacological activities and has fleshy juicy fruit with high nutritional value and is rich in vitamins. **Objectives:** This study aims to authenticate and examine the evolutionary relationships between various *Averrhoa* taxa. **Materials and Methods:** The species identification was verified using the well-established DNA barcoding method. For this, a few predetermined loci including rpoB, matK, atpF-atpH, and rbcL were employed. **Results:** Amplicons were observed with a molecular weight of 450bp, 950bp, 800bp, and 650bp respectively as reported earlier. The sequences of these loci were submitted to NCBI. The sequences were approved with accession numbers MN447319, MN733917, MT195537, and MT195538 respectively. The amplified loci of the species and related taxa of the genus were compared using bioinformatics tools. 204 numbers of SNPs/indels were identified out of which some are unique for the species. K2P analysis showed that the maximum intraspecific distance (1.3%) was lower than the minimum interspecific distance (2.6%). **Conclusion:** Among the different universal barcoding loci, matK was found to be most informative in terms of the occurrence of SNPs and genetic distance. So, the information can be utilized for species authentication in the family *Oxalidaceae*.

Keywords: *Averrhoa carambola,* DNA Barcoding, atpF–atpH, matK, rpoB.

Correspondence
Dr. Laxmikanta Acharya
Molecular Biology and Genetic Engineering Lab, Centre for Biotechnology, School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, INDIA.
Email id: laxmikantaacharya@soa.ac.in
ORCID ID: 0000-0002-8434-8500

Received: 17-06-2022;
Revised: 14-09-2022;
Accepted: 11-10-2022

INTRODUCTION

Averrhoa carambola L. is commonly known as the star fruit plant and it is included in the family *Oxalidaceae.* The distribution of this plant species is observed in most tropical countries like Malaysia, Indonesia, India, and the Philippines. This *Carambola* species shares the genus *Averrhoa w*ith other four different species namely *A. microphylla, A. bilimbi*, *A. leucopetala*, and *A. dolichocarp. Averrhoa carambola* is widely cultivated on a commercial scale mostly in countries like Southeast Asia and Malaysia.[1-3] The fleshy and juicy star fruits are naturally acidic, slightly tart, and sweet in test when ripening. The carambola tree has a short trunk of up to 30 feet in height holding many branches and leaves are green in color having a length of 6-11 inches with 5 to 11 ovate leaflets. The flowers are purple and are about 0.25 inches wide. Fruit is oval with a length of 2-6 inches, green in color but after ripening its color change to pale yellow. The fruit has five to six prominent longitudinal ridges and it looks like a star in cross-section. This fruit contains 10 to 12 numbers of flat light brown color seeds and they lose viability within a week if they are isolated from the fruit.

Averrhoa carambola contains anti-oxidant properties and can be used for the treatment of hypoglycemia and hypocholesterolemia.[4-6] Star fruit can be eaten raw but commonly it is consumed by making its juice, fruit salad, and delicious pickles. It is used for household purposes as it helps in cleaning utensils by removing rust. It is a natural source of antioxidants and a rich source of Vitamin C. in addition to that it contains various important phytochemicals and dietary fibers.[7]

It is laxative in nature and used as a traditional medicine to fight against fever, many skin diseases, high blood pressure, and diabetes. Different parts of these plants are having various functions to fight against various diseases as seeds are used against jaundice, and asthma, and fruits are analgesics and antioxidant agents, whereas leaves are used against hypertensive and ulcers. The stem is used as an anti-inflammatory, antimicrobial and anti-tumor agent. Previously from a study, it was found that this fruit was used for various treatments like throat pain, toothache,

DOI: 10.5530/ijper.57.2s.39

scabies, vomiting fever, and eye infection.[8,9] Though it has various pharmaceutical importance this plant species has to be identified and research should be carried forward for possible drug discoveries.

Species identification in plants is carried out by employing several techniques. Conventionally species identification is done by considering different morphological characteristics. But nowadays different molecular techniques are available for that purpose. DNA barcoding is one of the most accepted techniques for the identification and comparison of different plant species.[10] From the past investigation, it was found that there is a presence of seven barcoding regions in the plastid.[11-13] Four regions i.e. matK, rbcL, rpoB, and rpoC1 represent intragenic conserved loci, and the other three regions i.e. atpF–atpH, trnH–psbA, and psbK–psbI represent intergeneric conserved loci.[14,15] To avoid confusion in identifying *Averrhoa* species, the present study has been carried out by applying DNA barcoding tools and the identification of SNPs (single nucleotide polymorphism).

MATERIALS AND METHODS

Sample Collection

Fresh and healthy leaves of the sample of *Averrhoa carambola* (Figure 1) were collected from different parts of Odisha. Samples were washed with tap water and dried with tissue paper.

Extraction of DNA from *Averrhoa carambola* Leaves

Doyle and Doyle's technique was applied for the extraction of DNA from *Averrhoa carambola* leaves.[16] According to this method, 2g of fresh and healthy leaves were weighed and crushed with a mortar and pestle by adding a pinch of PVP (Polyvinylpyrrolidone) powder and 10 ml of extraction buffer containing 2% CTAB, 100mM Tris, 1.4M NaCl, 20 mM EDTA, 1% β-mercaptoethanol. The mixture was incubated in a water bath at 65°C for 1hr. After incubation for 1hr, it was cooled to room temperature, and an equal amount of chloroform: isoamyl alcohol (24:1) solution mixture was added to it and mixed well. Then the centrifugation was carried out at 10,000 rpm for 20 min at room temperature. After centrifugation, the supernatant was collected in a fresh microcentrifuge tube, and to it, a double amount of dehydrated ethanol was added for the precipitation of total DNA. The appearance of a white cottony mass of DNA was observed. This was taken out to another tube and washed three times with 70% ethanol and left for drying. Resuspension of DNA was carried done by adding 1ml of $T_{10}E_1$ (Tris-EDTA) buffer. For removing RNA, RNase treatment was carried out by adding RNase A in a concentration of 5µg/µl and incubated at 37°C for 1 hr in a water bath. Again chloroform: isoamyl (24:1) treatment was given to purify DNA followed by centrifugation and precipitation process. Then DNA was diluted with $T_{10}E_1$ buffer and quantified against λ-DNA as standard and further diluted for working concentration to 25ng/µl.

Figure 1: *Averrhoa carambola* (with flower and fruit) in its natural habitat.

Polymerase Chain Reaction (DNA amplification)

A few sets of barcoding primers (Table 1) were used to amplify the conserved regions. PCR was carried out by preparing 25µl of the master mix containing template DNA (25ng), 10X PCR buffer, 2.5mM dNTP mixture, 1µl forward and reverse primer, 1U Taq DNA Polymerase, and Nuclease free water. The DNA was amplified in Veriti™ 96-Well thermocycler (Applied Biosystems™) by taking the following parameters: 96°C for 5min for initial denaturation, 92°C for 1min for denaturation, subsequent primer annealing for 1min at primer-specific temperature, and extension 72°C for 2 min. The process was repeated for 35 cycles with a subsequent final extension at 72°C for 7 min. Amplified

Table 1: Primer sets used for PCR amplification.

Barcoding regions	Sequences	Primer name
rpoC1	GGCAAAGAGGGAAGATTTCG	LKB1
	CCATAAGCATATCTTGAGTTGG	LKB2
rpoB	ATGCAACGTCAAGCAGTTCC	LKB3
	CCGTATGTGAAAAGAAGTATA	LKB4
matK	CGTACAGTACTTTTGTGTTTACGAG	LKB5
	ACCCAGTCCATCTGGAAATCTTGGTTC	LKB6
	TCTAGCACACGAAAGTCGAAGT	LKB7
	CGATCTATTCATTCAATATTTC	LKB8
	GTTCTAGCACAAGAAAGTCG	LKB9
	TAATTTACGATCAATTCATTC	LKB10
atpF-atpH	ACTCGCACACACTCCCTTTCC	LKB11
	GCTTTTATGGAAGCTTTAACAAT	LKB12
psbK-psbI	TTAGCCTTTGTTTGGCAAG	LKB13
	AGAGTTTGAGAGTAAGCAT	LKB14
rbcL	GTAAAATCAAGTCCACCRCG	LKB15
	ATGTCACCACAAACAGAGACTAAAGC	LKB16
trnH-psbA	GTTATGCATGAACGTAATGCTC	LKB17
	CGCGCATGGTGGATTCACAATCC	LKB18

PCR products were resolved in a gel composed of 1.5% agarose added with ethidium bromide. Then the DNA fragments were eluted using Silica Bead DNA Gel Extraction Kit-Thermo Fisher Scientific and were reamplified by using the same primer. Then the samples were sequenced by the Sanger sequencing method and validated in the BLAST search tool.

Phylogenetic Relationship

To determine the phylogenetic relationship from the BLAST analysis result between various plant species MEGA 10 was used. The sequence of resulting barcoding regions i.e. rpoB, matK, atpF–atpH, and rbcL of ten various plant species were aligned by the MUSCLE program and the Phylogenetic tree was constructed by the neighbor-joining method in the MEGA X program and the distance between the inter and intra population was evaluated from the phylogenetic analysis.[17,18]

RESULTS

Resolving Barcoding Regions and Sequence Submission

Seven conserved nuclear loci and chloroplast DNA loci were taken into consideration for DNA barcoding to identify *Averrhoa carambola*. Four of the seven loci generated different and repeatable amplifications. Amplicons were resolved with 450bp, 950bp, 800bp, and 650bp for the rpoB, matK, atpF-atpH, and rbcL loci, respectively (Figure 2). The sequences for these loci were deposited at NCBI. Following examination, NCBI has given the aforementioned areas the accession numbers MN447319, MN733917, MT195537, and MT195538, respectively.

Detection of SNPs and In/Dels

The current sequencing findings were compared to similar sequence data of several plant species included in the public domain database (NCBI). The query coverage for the top ten sequences was estimated to be between 90% and 100%. The Sequencher software was used to detect SNPs and In/Dels, and an analysis of the total number of SNPs, their range, and In/Del location was performed. The genus is confirmed by BLAST analysis of the conserved loci, and variations in the sequence demonstrate the specificity of the loci for the species. Multiple sequence alignment using the MUSCLE tool was used to compare the rpoB sequence of *Averrhoa carambola* with that of the species and a few other species (Figure 3), and a total of 107 SNPs were discovered that could be utilized for species identification (Table 2). In an average length of around 4 base pairs, one SNP can be found. Along with carambola species, the matK sequence was also compared with other species (Figure S1), and it was discovered that 204 SNPs may be useful for species identification (Table S1). One of these 204 SNPs is unique to other Carambola species. An average length of about three base pairs can include one SNP. Once again, the atpF-atpH sequence of *Averrhoa carambola* was compared with various species in addition to Carambola species (Figure S2), and it produced 93 numbers of SNPs, out of which four may be utilized for additional Carambola species identification (Table S2). In this instance, a single SNP may be seen in a length of around 5 base pairs on average. The rbcL sequence of *Averrhoa carambola* had a total of 23 SNPs, of which three could be used for identification (Table S3), and one covered an average length of 23 base pairs (Figure S3). The sequences were initially aligned with a CLUSTAL algorithm called MUSCLE 3.8 to determine the location of In/Dels.[19] Table 3 lists the locations of In/Dels that were discovered in the rpoB, matK, atpF-atpH, and rbcL sequences. There were no such In/Del sites detected in the atpF-atpH and rbcL loci, however, a total of 11 numbers of In/Del positions were acquired from the rpoB region, and 18 numbers of In/Del positions were obtained from the matK region.

As a result, the rpoB and matK In/Dels and SNPs that were acquired might be used as a qualitative molecular marker for species authentication.[20-24]

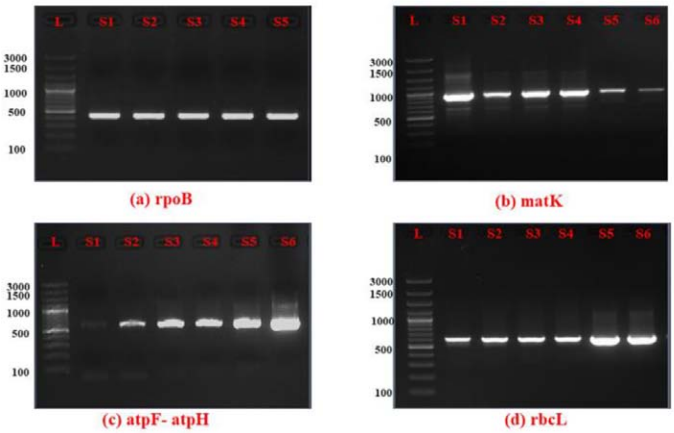

Figure 2: DNA amplification profile of different loci.

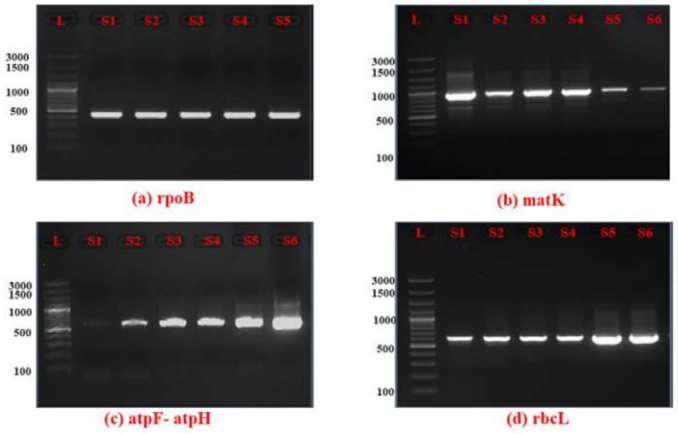

Figure 3: Multiple sequence alignment showing SNPs and In/Dels positions in rpoB loci.

Table 2: SNP variance table for rpoB loci.

Reference	KX364202.1 *Averrhoa carambola*	MN447319.1 CBT *Averrhoa carambola*	NC_048890.1 *Oxalis corymbosa*	MW190090.1 *Sloanea sinensis*	MT985378.1 *Elaeocarpus japonicus*	MT593043.1 *Elaeocarpus braceanus*	MN615725.1 *Brunellia antioquensis*	MN585217.1 *Brunellia trianae*	KU569488.1 *Averrhoa carambola*	JX663460.1 *Averrhoa carambola*	KX364202.1 *Averrhoa carambola*	Total
25	T						C	C				2
37	C				T	T						2
40	G		A									1
44	C						T	T				2
52	T		C									1
69	A			G	G	G	G	G				5
77	C		A	A	A	A	A	A				6
80	T		C									1
107	A			G	G	G	G	G				5
128	G		A									1
131	C		A	G	A	A	A	A				6
152	C		T									1
185	T			C	C	C	C	C				5
192	A			G								1
193	C				T	T						2
241	T				C	C						2
248	T		C		C	C						3
257	T		A									1
269	G				:	:						2
272	T			:			:	:				3
273	T			G	G	G	G	G				5
276	A		:	:	:	:	:	:				6
277	A		:	:	:	:	:	:				6
278	C		:									1
279	A		G	G	G	G	G	G				6
286	T		C	C	C	C	C	C				6
287	G						A	A				2
304	T				C							1
315	C		T	T	T	T	T	T				6
320	C		T	T								2
339	G		A									1
357	G						A	A				2
362	T				G	G						2
374	T		C									1
380	C		T	T	T	T	T	T	T			7
425	T			C								1
Total		0	19	16	19	18	17	17	1	0	0	107

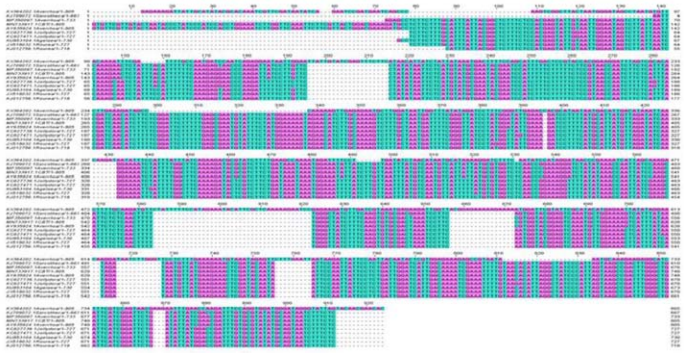

Figure S1: Multiple sequence alignment showing SNPs and In/Dels positions in matK loci.

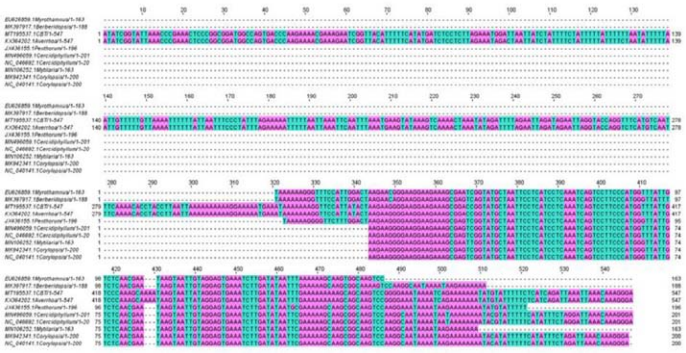

Figure S2: Multiple sequence alignment showing SNPs and In/Dels positions in atpF-atpH loci.

Table S1:SNP variance table for matK loci.

Reference	KX364202.1 *Averrhoa carambola*	KJ709072.1 *Sarcotheca griffithii*	KJ012756.1 *Rourea surinamensis*	KC627471.1 *Jollydora duparquetiana*	KC627739.1 *Jollydora glandulosa*	JX518032.1 *Rourea orientalis*	KU853104.1 *Agelaea macrophylla*	MF350097.1 *Averrhoa carambola*	MN733917.1 CBT *Averrhoa carambola*	AY935924.1 *Averrhoa carambola*	KX364202.1 *Averrhoa carambola*	Total
12	T		C	C	C	C	C					5
32	T	G	G	G	G	G	G					6
35	G		C	C	C	C	C					5
66	G		A	A	A	A	A					5
95	G				C							1
115	C		T	T	T	T	T					5
130	G		A									1
175	C	T										1
179	T		G			G	G					3
186	G		A			A	A					3
188	C		G	G	G	G	G					5
193	C						T					1
196	G		A	A	A	A	A					5
208	G						A					1
214	C		T	T	T	T	T					5
215	G	T	T	T	T	T	T					6
225	A	C										1
245	G		A	A	A	A	A					5
261	G		A	A	A	A	A					5
266	C					T						1
269	G		C	C	C	C	C					5
287	G		A	A	A	A	A					5
289	T	C	A	A	A	A	A					6
296	G	C	C	C	C	C	C					6
309	G		A	A	A	A	A					5
312	T		G									1

continued...

Table S1: Cont'd.

Reference	KX364202.1 *Averrhoa carambola*	KJ709072.1 *Sarcotheca griffithii*	KJ012756.1 *Rourea surinamensis*	KC627471.1 *Jollydora duparquetiana*	KC627739.1 *Jollydora glandulosa*	JX518032.1 *Rourea orientalis*	KU853104.1 *Agelaea macrophylla*	MF350097.1 *Averrhoa carambola*	MN733917.1 CBT *Averrhoa carambola*	AY935924.1 *Averrhoa carambola*	KX364202.1 *Averrhoa carambola*	Total
315	T		A	A	A	A	A					5
319	G		A	A	A	A	A					5
345	T		A	A	A	A	A					5
371	C		A	A	A	A	A					5
387	A								G			1
392	A						C					1
393	G		A			A	A					3
422	A		G	G	G	G	G					5
436	A				C							1
439	T		A	A	A	A	A					5
442	A					C						1
452	C	T										1
453	C					T						1
466	C		T			T	T					3
490	T		C			C	C					3
528	G		A	A	A	A	A					5
531	G			A	A							2
545	A			C	C							2
549	G			A	A							2
563	G	A										1
575	A		G	G	G	G	G					5
531	G			A	A							2
545	A			C	C							2
549	G			A	A							2
563	G	A										1
575	A		G	G	G	G	G					5
584	G			A	A							2
590	G		A									1
614	C			G	G							2
618	G		A	A	A	A	A					5
620	T		A	A	A	A	A					5
632	C	A	A			A	A					4
645	A			G								1
648	T		G			G	G					3
656	G		T	T	T	T	T					5
658	T		G	G	G	G	G					5
663	G			A	A							2
664	A		G	G	G	G	G					5
686	A	X		C	C							2
695	G	X	A									1
731	A	X	X	X	X	X	X	:				1
Total	9	40	37	38	39	39	1	1	0	0	204	Total

Table S2: SNP variance table for atpF-atpH loci.

Reference KX364202.1 *Averrhoa carambola*		MT195537.1 CBT *Averrhoa carambola*	KX364202.1 *Averrhoa carambola*	EU626859.1 *Myrothamnus flabellifolia*	MK397917.1 *Berberidopsis beckleri*	JX436155.1 *Penthorum chinense*	MN106252.1 *Mytilaria laosensis*	Total
97	A	G		X	X	X	X	1
100	C	T		X	X	X	X	1
128	C	T		X	X	X	X	1
268	C	T		X	X	X	X	1
328	A					G	X	1
329	A			G		G	X	2
333.1	:			T	T		X	2
335	C					:	X	1
336	A					T	X	1
339	A			G	G	G	X	3
340	T			G	G	G	X	3
349	G			C	C			2
350	G				A			1
366	G			A	A	A	A	4
367	A				G	G		2
368	C			T	T	T	T	4
369	C						T	1
370	G				A			1
411	T				G	G	G	3
417	G				T			1
420	C			T	T	T	T	4
424	A			:	:	:	:	4
425	G			:	:	:	:	4
427	A			:	:	:	:	4
428	A			G	G	G	G	4
440	A			G	G	G	G	4
448	G					A		1
464	T					G		1
465	T					C	C	2
466	G			A			A	2
472	A			G	G	G	G	4
473	A			C	C	C	C	4
477	C			T				1
478	A			G	G	G	G	4
482.1	:				A			1
482.2	:				A			1
487	G			X	A	A	A	3
488	G			X	A	A	A	3
491	A			X	C		C	2
500	C			X	A	A	A	3
509.1	:			X	X	A	X	1
Total		4	0	18	26	26	19	93

Table S3: SNP variance table for rbcL loci.

Reference	KX364202.1 *Averrhoa carambola*	MT195538.1 CBT *Averrhoa carambola*	MG833489.1 *Connarus regnellii*	MG718075.1 *Connarus rostratus*	GQ436319.1 *Averrhoa carambola*	KX364202.1 *Averrhoa carambola*	FJ670180.1 *Averrhoa carambola*	MF349621.1 *Averrhoa carambola*	JX572418.1 *Cnestis polyphylla*	AY935743.1 *Averrhoa carambola*	AY788196.1 *Dapania racemosa*	Total
48	C										A	1
139	C		G	G					G			3
196	T								C			1
232	C		T	T								2
265	T		C	C					C			3
277	T		G	G					G			3
478	A		G	G					G		C	4
529	G	A										1
536	G	T										1
540	G	T										1
544	A		G	G					G			3
Total		3	6	6	0	0	0	0	6	0	2	23

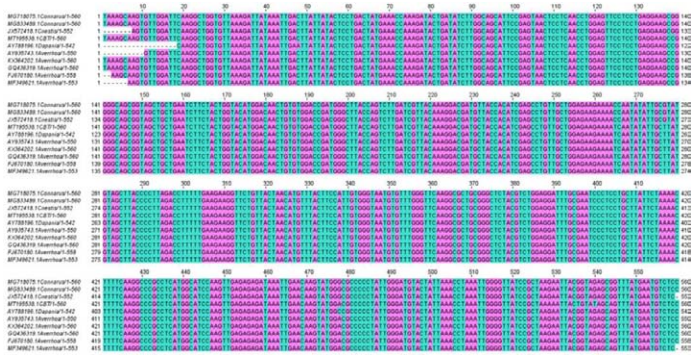

Figure S3: Multiple sequence alignment showing SNPs and In/Dels positions in rbcL loci.

Phylogenetic Analysis

The *Brunelliaceae* family is connected to *Elaeocarpaceae*, while *Elaeocarpaceae* and *Oxalidaceae* form a separate cluster, according to phylogenetic analysis of the rpoB sequence from diverse plant species. With an 87% resemblance to previously reported *Averrhoa carambola*, CBT *Averrhoa carambola* was the most closely related, and it constitutes a new group in Figure 4. Here, the genetic distance between and among species was assessed. It was discovered that the minimum interspecific genetic distance was 0.026% and the greatest intraspecific genetic distance was 0.013%.

Two separate clusters were found in the phylogeny of the matK regions of different plant species; one cluster comprised all the

Table 3: In/Dels positions of various loci.

Sl. No	Barcoding region	In/Dels position
1	rpoB	1-24
2		53-57
3		59
4		115-116
5		141-143
6		156
7		323-325
8		354-355
9		412-433
10		453-457
11		466-470
1	matK	1-12
2		57-59
3		79-108
4		155-160
5		196-215
6		299-337
7		395
8		427-431
9		491-496
10		531
11		584-622
12		656-671
13		710-711
14		716-726
15		756-764
16		812-833
17		868-870
18		914-924
1	atpF-atpH	NO In or Del
1	rbcL	NO In or Del

species belonging to the *Connaraceae* family, while the other cluster contained all the species belonging to the Oxalidaceae family. Maximum similarity, about 85%, between CBT *Averrhoa carambola* and preceding carambola species (Figure 5). In this case, the minimum interspecific genetic distance was 0.03% and the greatest intraspecific genetic distance was 0.015%.

The families *Hamamelidaceae*, *Penthoraceae*, and *Cercidiphyllaceae* were shown to belong to a single cluster in the phylogenetic trees constructed from the atpf-atpH regions of different plant

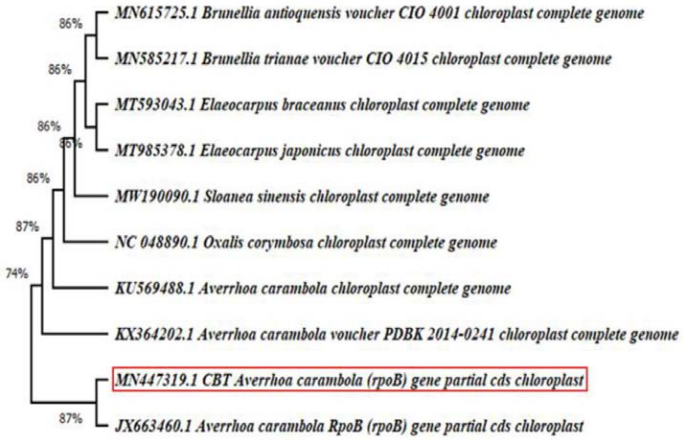

Figure 4: Phylogenetic tree based on rpoB sequences of various plant species.

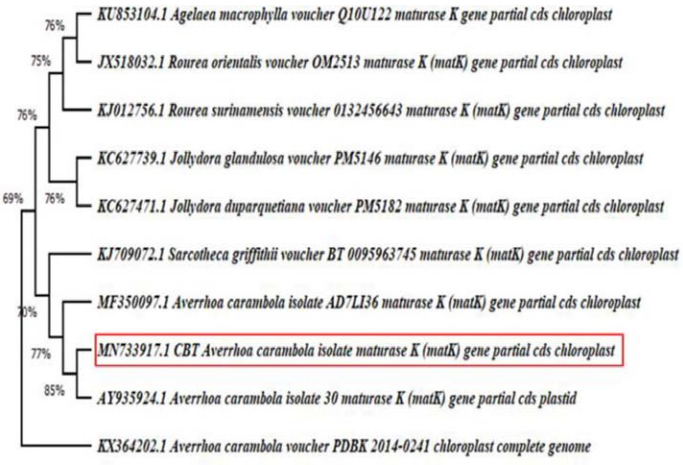

Figure 5: Phylogenetic tree based on matK sequences of various plant species.

species, but the families *Berberidopsidaceae*, *Myrothamnaceae*, and *Oxalidaceae* are distantly related. CBT *Averrhoa carambola* developed a distinct group and shares 99% similarity with earlier carambola species, as shown in Figure 6. A maximum intraspecific genetic distance of 0.01% and a minimum interspecific genetic distance of 0.029% were identified in this case.

In the Oxalidaceae family and the *Connaraceae* family, the rbcL loci of various plants also formed distinct groupings. Figure 7 from the CBT *Averrhoa carambola* revealed that the two species were 100% identical. The maximum intraspecific genetic distance, in this case, was determined to be 0.001%, while the lowest interspecific genetic distance was determined to be 0.003%.

DISCUSSION

Plant DNA barcoding is important because accurate species identification is required to preserve and utilize plants. However, due to a lack of taxonomic knowledge in many parts of the world, this may be hindered.[25] Parts of two plastid genes, rbcL and matK,

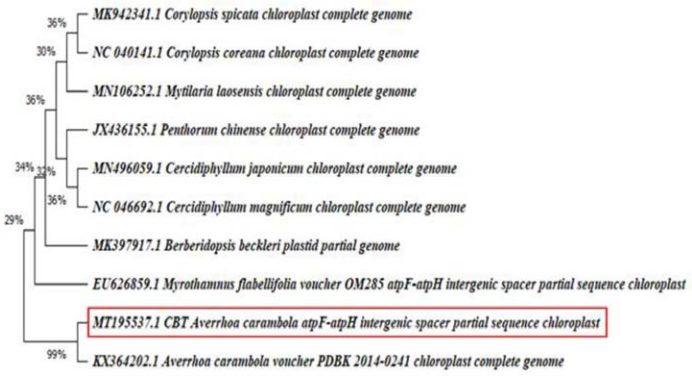

Figure 6: Phylogenetic tree based on atpF-atpH sequences of various plant species.

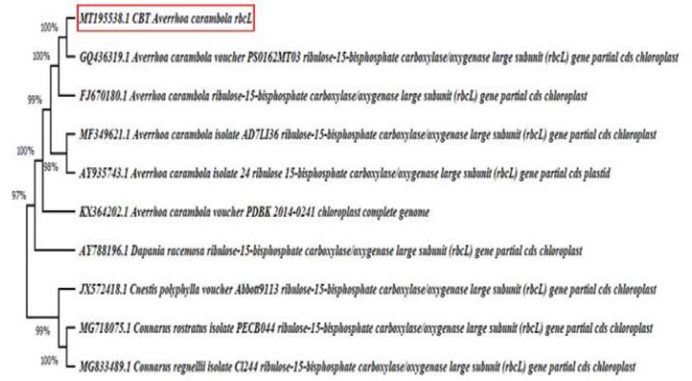

Figure 7: Phylogenetic tree based on rbcL sequences of various plant species.

were chosen as the standard plant DNA barcodes after the Plant Working Group (PWG) of the Consortium for the Barcoding of Life (CBOL) reviewed a variety of potential markers, with the awareness that more markers may be required.[26]

Four of the seven barcode loci chosen for this investigation were able to be resolved, and the acquired sequences corresponded with the previously available data for the same species, allowing the species to be recognized by NCBI GenBank.

An SNP only modifies one nucleotide, but an indel affects one or more nucleotides in the DNA sequence.[27] It was discovered that different SNPs and Indels were present in all of the amplified sites. These areas could be used as species identification markers.

For use in phylogenetic community ecology, phylogenetic trees have also been created using genetic sequences supplied by DNA barcoding.[28] The use of DNA barcoding as an identifying technique also requires the creation of excellent reference databases for sequences.[29] From the phylogenetic analysis, the inter and intra-specific distance was revealed which indicates the maximum intraspecific distance (1.3%) was lower than the minimum interspecific distance (2.6%). A DNA sequence is regarded as the ideal barcode identifier for species differentiation if it exhibits higher interspecific and lower intra-specific divergence.[18]

CONCLUSION

Molecular identification of *Averrhoa carambola* was carried out using DNA barcoding markers. It is quite challenging to distinguish between cryptic species of *Averrhoa* throughout the vegetative and early developmental stages. So, in the present investigation authentication of *Averrhoa carambola* has been done and a phylogenetic relationship among related species is established using DNA sequence data revealed from conserved barcode regions. Unique species-specific SNPs detected for this species can be used as a tool for authentic identification of the taxon. It also helps in finding an appropriate phylogenetic relationship among the studied taxa.

ACKNOWLEDGEMENT

The President and Vice-Chancellor of SOA (Deemed to be University) are highly acknowledged for the infrastructure facility.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

DNA: Deoxy Ribonucleic Acid; **RNA:** Ribonucleic acid; **rpoB:** RNA polymerase β subunit; **matK:** Megakaryocyte-Associated Tyrosine Kinase; **rbcL:** RuBisCO large subunit; **NCBI:** National Center for Biotechnology Information; **PVP:** Polyvinylpyrrolidone; **EDTA:** Ethylenediaminetetraacetic acid; **CTAB:** cetyltrimethylammonium bromide; **PCR:** Polymerase Chain Reaction; **BLAST:** Basic Local Alignment Search Tool; **SNP:** Single Nucleotide Polymorphism; **In/Del:** Insertion/Deletion.

SUMMARY

Averrhoa carambola L. (star fruit) is an underutilized plant under the family Oxalidaceae. The plant has fleshy juicy fruit with high nutritional value and is rich in vitamins. There is confusion in the identification of the species for non-taxonomists in the vegetative stage and herbal formulations. Here universally accepted DNA barcoding technique was used to authenticate the species identification. Some prescribed loci namely rpoB, matK, atpF–atpH, and rbcL were amplified and sequence loci were submitted at NCBI. The amplified loci of the species and related taxa of the genus were compared and 204 numbers of SNPs/indels were identified out of which some are unique for the species. K2P analysis showed that the maximum intraspecific distance (1.3%) was lower than the minimum interspecific distance (2.6%). Among the different universal barcoding loci, matK was found to be most informative in terms of the occurrence of SNPs and genetic distance. So, the information can be utilized for species authentication in the family *Oxalidaceae*.

REFERENCES

1. Khoo HE, Azlan A, Kong KW, Ismail A. Phytochemicals and medicinal properties of indigenous tropical fruits with potential for commercial development. Evid Based Complement Alternat Med. 2016;2016:7591951. doi: 10.1155/2016/7591951, PMID 27340420.
2. HockEng K, Prasad KN, KinWeng K, LyeYee C, Azlan A, Jian S, *et al*. A review on underutilized tropical fruits in Malaysia. Guangxi Agric Sci. 2010;41(7):698-702.
3. JF M. Fruits of warm climates. Echo point books media. 2013.
4. Chau CF, Chen CH, Lin CY. Insoluble fiber-rich fractions derived from Averrhoa carambola: Hypoglycemic effects determined by *in vitro* methods. LWT Food Sci Technol. 2004;37(3):331-5. doi: 10.1016/j.lwt.2003.10.001.
5. Chau CF, Chen CH, Wang YT. Effects of a novel pomace fiber on lipid and cholesterol metabolism in the hamster. Nutr Res. 2004;24(5):337-45. doi: 10.1016/j.nutres.2004.01.003.
6. Shui G, Leong LP. Residue from star fruit as valuable source for functional food ingredients and antioxidant nutraceuticals. Food Chem. 2006;97(2):277-84. doi: 10.1016/j.foodchem.2005.03.048.
7. Ayala-Zavala JF, Vega-Vega V, Rosas-Domínguez C, Palafox-Carlos H, Villa-Rodriguez JA, Siddiqui MW, *et al*. Agro-industrial potential of exotic fruit byproducts as a source of food additives. Food Res Int. 2011;44(7):1866-74. doi: 10.1016/j.foodres.2011.02.021.
8. Payal G, Pankti K, Manodeep C, V KJ. Phytochemical and pharmacological profile of Averrhoa carambola Linn: An overview. Int Res J Pharm. 2012;3(1):88-92.
9. Sheth A, Mitaliya KD, Joshi S. The herbs of Ayurveda. 1st ed. Gujarat, India: Sheth Publisher; 2005.
10. Asif H, Khan A, Iqbal A, Khan IA, Heinze B, Azim MK. The chloroplast genome sequence of *Syzygium cumini* (L.) and its relationship with other angiosperms. Tree Genet Genomes. 2013;9(3):867-77. doi: 10.1007/s11295-013-0604-1.
11. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A. 2005;102(23):8369-74. doi: 10.1073/pnas.0503123102, PMID 15928076.
12. Chase MW, Cowan RS, Hollingsworth PM, Van Den Berg C, Madriñán S, Petersen G, *et al*. New trends in plantsystematics a proposal for a standardised protocol to bar code all land plants. Nat Hist. 2007;56(May):295-9.
13. Ford CS, Ayres KL, Toomey N, Haider N, Van Alphen Stahl J, Kelly LJ, *et al*. Selection of candidate coding DNA barcoding regions for use on land plants. Bot J Linn Soc. 2009;159(1):1-11. doi: 10.1111/j.1095-8339.2008.00938.x.
14. Pennisi E. Taxonomy.wanted: A bar code for plants. Science. 2007;318(5848):190-1. doi: 10.1126/science.318.5848.190, PMID 17932267.
15. Ledford H. Botanical identities. Nature. 2008;451(7179):616. doi: 10.1038/451616b, PMID 18256630.
16. Doyle JJ, Doyle J. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13-5.
17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547-9. doi: 10.1093/molbev/msy096, PMID 29722887.
18. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004;101(30):11030-5. doi: 10.1073/pnas.0404206101, PMID 15258291.
19. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25(24):4876-82. doi: 10.1093/nar/25.24.4876, PMID 9396791.
20. Rebecca LJ, Kumar SA, Sharmila D, Kowsalya E, Lakshmi RMG. Evaluation of Genetic Diversity among *Ocimum sanctum*. Res J Pharm Technol. 2019;12(6):2889. doi: 10.5958/0974-360X.2019.00487.6.
21. Sheng T, Lu Y, Yang K, Jin Y, Wu Y, Zhang Z, *et al*. Association between single nucleotide polymorphisms (SNPs) in the promoter of adiponectin gene, hypoadiponectinemia, and diabetes. Res J Sci Technol. 2016;8(1):34. doi: 10.5958/2349-2988.2016.00004.8.
22. Al-Koofee DAF, Mobarak SMH. PRIMER1: A network service for tetra-arms PCR primer design based on well-known DbSNP. Res J Pharm Technol. 2018;11(8):3633-7. doi: 10.5958/0974-360X.2018.00669.8.
23. Singh A, Singh S, Anbarasu A. *In silico* evaluation of non-synonymous snps in irs-1 gene associated with type II diabetes mellitus. Res J Pharm Technol. 2018;11(5):1957-61. doi: 10.5958/0974-360X.2018.00363.3.
24. Anbarasu K, Jagadeeswari P, Mahendran R. *In-silico* screening of deleterious NF1 SNPs associated with neurofibromatosis Type I. Res J Pharm Technol. 2017;10(9):3247. doi: 10.5958/0974-360X.2017.00576.5.
25. Chase MW, Fay MF. Ecology. Barcoding of plants and fungi. Science. 2009;325(5941):682-3. doi: 10.1126/science.1176906, PMID 19644072.
26. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci U S A. 2009;106(31):12794-7. doi: 10.1073/pnas.0905845106, PMID 19666622.
27. Loewe L. Genetic mutation. Nature education. 2008;1:1(1):113.
28. Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, *et al*. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci U S A. 2009;106(44):18621-6. doi: 10.1073/pnas.0909820106, PMID 19841276.
29. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLOS ONE. 2011;6(5):e19254-. doi: 10.1371/journal.pone.0019254, PMID 21637336.

Cite this article: Giri AK, Biswal B, Jena B, Acharya L. Authentication of *Averrhoa carambola* L. Using DNA Barcoding: Detection of SNPs and Evaluation of Phylogenetic Relationship. Indian J of Pharmaceutical Education and Research. 2023;57(2s):s335-s344.