International Journal of Pharmaceutical Science Invention
ISSN (Online): 2319-6718, ISSN (Print): 2319- 670X
www.ijpsi.org Volume 9 Issue 6 ǁ June 2020 ǁ PP. 24-34

Development of DNA barcodes for Medicinal plant *Sida* species in Sri Lanka

R. Rajphriyadharshini[1], O.V.D.S.J Weerasena[1], T.M.S.G. Tennakoon[2], and L.N.R Gunaratna[1]

[1]*Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Sri Lanka.*
[2]*Link Natural Products (PVT) Ltd, Sri Lanka*
Corresponding Author:O.V.D.S.J Weerasena

ABSTRACT: *Babilla (Sida) is known for its medicinal value including in vitro cytotoxic activity. It has enormous applications in the Ayurvedic system of medicine. Sida species are substituted by other plants of the same genus in the herbal medicinal industry due to morphological similarity. Additionally, taxonomists find it difficult to identify the authentic Sida plant from processed products. Thus this obstacle can now be overcome by using DNA barcoding. Due to its economical and biological significance developing a DNA barcode may contribute to the trade control and conservation of Sida species. The general objective of this study is to develop barcodes to discriminate Sri Lankan Sida species. As DNA barcodes, four chloroplast DNA regions which were proposed by previous studies in plant barcoding, known as matK. rbcL, ITS-2, and psbA-trnH were selected. The barcode regions were amplified and sequenced. Phylogenetic trees were constructed using MEGA V 7. The results suggested that DNA barcodes have the potential to identify Sida species. Based on the intra-species divergence observed, the ITS-2 and rbcL were found to be the most suitable regions to distinguish Sida species out of the four barcode regions. The rbcl barcode region is ambiguous in assigning Sida species into the correct family or genus, the barcode sequences resulted in this study were made available in the gene bank database. The current study strongly suggests that the raw drug market samples of herbal medicines need to be properly authenticated before use, and DNA barcoding is suitable for this purpose.*

KEYWORDS: *DNA, Sida, Barcoding, Adulteration, Sequencing*

Date of Submission: 20-08-2020 Date of acceptance: 15-09-2020

I. INTRODUCTION

Sida is a genus of flowering plants of Malvaceae that includes about 200 species distributed in tropical and subtropical regions of the world. *Sida* showed a wide range of morphological variability. Nearly 18 species of *Sida* (*Sidae, Sida Lilm., Nelvagae, Cordifoliae, Spinosae, Malachroideae,* and *Stenindae,* etc) were known to occur in Sri Lanka (Aguilar et al. 2003). Even though so far 200 species were recorded under the *Sida* genus, over 1000 species were predicted under the genus, due to uncertain morphological variability of *Sida* plants. In contrast, some studies including Santhosh et al., (2015), Simplice et al., (2007), andVassou et al., (2015) accepted that *Sida* only includes 98 species in it. Thus the diversity of *Sida* is still unclear which gives rise to a high rate of adulteration of *Sida* plants.

Adulteration usually refers to mixing other matters of an inferior which may occasionally harmful to the quality of food, drink, or chemical product that are intended to be sold. Adulteration of herbal products includes admixtures, product substitution, and contamination which could causelife-threatening health implications. Thus it is important to develop an accurate method that confirms the authenticity of raw material, which is a fundamental component of quality assurance in the field of the food supply industry. Currently, there were many chemical fingerprints developed to unveil the authenticity of raw material and to overcome species adulteration. In herbal plants, drug adulteration usually refers to mixing completely different plants with genuine plant materials. Substitution refers to replace one genuine plant by other plants of the same genus. Hence DNA barcoding help to establish the correct authentication of the plants that undergo both adulteration and substitution. *Sida* is a medicinal plant that has numerous medicinal applications. Many of its biological activities, such as cytotoxicity, anti-inflammatory, antibacterial, antifungal, antiplasmodial, and antioxidant are revealed. This plant is famous as a constituent of Ayurvedic formulations to treat the neuronic disorder. However, *Sida* is substituted by other plants of the same genus in trade, which reduces its beneficial effects to the public. Further, this valuable medicinal plant needs conservation. Barcoding allows providing proper identification of the plant. Moreover, DNA barcodes of this species will be useful to monitor the adulteration of other herbs and produce a successful tool for the plant recognition up to subspecific taxa. The objectives of this study are to develop barcodes for authentic *Sida* plants that are commonly used in traditional medicine and

to decide which locus out of matK, psbA-trnH, ITS-2, and rbcL is more suitable for easy barcoding of *Sida* species.

.

II. MATERIALS AND METHOD

2.1 DNA Extraction

DNA was extracted using CTAB based plant DNA mini-preparation method described own and Ronald (1999) with several modifications. Sida young leaves were washed with deionized water and wiped with 70°C ethanol. Leaves were placed 2ml Eppendorf tube. Blunted 1000 µl tips were used to grind the leaf tissues to a line powder in liquid nitrogen. The samples were immediately placed in the ice. CTAB Extraction buffer (see Appendix 1) was preheated to 65 ºC and 700 µl were added to each Eppendorf tube. To prevent RNA contamination 7 µl of 20 µg/ml RNase was added into the mixture. The samples were mixed thoroughly. The mixture was incubated at 65ºC in a water bath for 8 minutes with occasional mixing. Then 570 µl of CHISAM (see appendix 1) was added to each sample and mixed well by inverting the samples. The samples were centrifuged at 13000 rpm at room temperature for 10 minutes, using a benchtop microcentrifuge (Eppendorf 580 R). The aqueous phase in the supernatant was transferred into a new Eppendorf tube without disturbing the interphase. DNA was precipitated by adding 0.7 volumes of Isopropanol into the aqueous layer containing tube and mixed gently for 15 minutes on ice. Then the mixture centrifuged at 13000 rpm at room temperature for 5 minutes. The supernatant was removed using a pipette. The pellet was rinsed with 150 µl of 70% ethanol. To remove ethanol the pellet was vacuum dried using a speed vac concentrator (Thermo savent ISSIIO speed vac system). The pellet was dissolved in 20 µl of autoclaved ultrapure water, followed by a brief spin and stored at -20ºC.

2.2 Agarose gel electrophoresis

Agarose gel was prepared (appendix 1) mixing agarose powder with 1xTris acetate electrophoresis (TAE) buffer and melted agarose by heating in a microwave oven. 0.5 µg/ ml Ethidium bromide was added into a melted agarose gel. The agarose gel solution was poured into a gel casting tray with a comb placed on a leveled surface. After solidification of agarose, the gel was transferred into the 1 x TAE buffer prepared from 50 x TAE stock (appendix 1) in a gel electrophoresis tank. The samples were loaded into wells and run at a constant voltage of 50 V (Electrophoresis power supply, EPS 301. Amersham Bioscience).

2.3 DNA quantification and quality evaluation

The concentrations of the DNA were determined by measuring the absorbance of each DNA sample at 260 nm and 280 nm wavelengths in a UV-visible spectrophotometer. Concentrations of the DNA samples were calculated and the purity of the samples was detected.

2.4 Polymerase Chain Reaction (PCR)

Primer sequences for the two coding regions matK, rbcL, and the noncoding regions of PSB-trnH and ITS-2 were obtained from the Consortium for the Barcoding of Life (CBOL) and the reference is given in Table 2.1.

Table 2.1: Oligonucleotide sequence of DNA Barcoding Primers used

Primer locus	Primer Name	Nucleotide sequence 5'-3'	Reference
rbcL	rbcLaF (forward)	5'-ATGTCACCACAAACAGAGACTAAAGC-3'	Levin,2003
	rbcLajf634R (reverse)	5'-GAAACGGTCTCTCCAACGCAT-3'	Kress and Erickson,2007
matK	matK : 3F_KIM (forward)	5'-CGTACAGTACTTTTGTGTTTACGAG-3'	Vassou et al,2015
	matK:1R_KIM (reverse)	5'-ACCCAG TCCATCTGGAAATCTTGGTTC-3'	Vassou et al,2015
psbA-trnH	psbA3'f (forward)	5'-GTTATGCATGAACGTAATGCTC-3'	Sang et al,1997
	trnH (reverse)	5'-CGCGCATGGTGGATTCACAATCC-3'	Tate and Simpson,2003
ITS2	ITS2: S2F (forward)	5'-ATGCGATACTTGGTGTGAAT-3'	Vassou et al,2015
	ITS2: S3R (reverse)	5'-GACGCTTCTCCAGACTACAAT-3'	Vassou et al,2015

2.5 Reconstitution of Primers

The stocks of both forward and reverse primers existed in the form of a powder. To prepare a working concentration 52.8 µl of PCR grade water was added to 26.4 nmol of the forward primer of rbcL as per the specifications in the product information sheet to yield a concentration of 500 µM. A brief pulse was given to collect the components to the bottom of the tube. The tubes were left at room temperature for 30 minutes to dissolve the content. Then a 10 µM solution of primer was prepared using the primer stock solution. PCR grade water 72.6 µl was added to the 36.3 nmol of reverse primer of rbcl to prepare a 500 µM solution. Similarly, 55 µl PCR water to psbA f 77.5 nmol, 74.6 µl PCR water to PSB-trnH. 37.3 nmol, 69.6 µl PCR water to matK 390f 34. 8 nmol, and 60.2 u1 PCR water to matK 136r 30.1 nmol were added. The same procedure of the rbcL forward primer preparation was applied to obtain the other 10 primer solutions. All primers were stored at -20°C.

2.6 Preparation of dNTP mixture

The working concentration of dNTP mixture (10 mM) was prepared by adding 30 µl PCR grade water to an Eppendorf tube with 5 µl of each 100 mM dNTP stock solution. Aliquots of working solutions were stored at -20 °C.

2.7 PCR Amplification

PCR was carried out using genomic DNA extracted from the dry power of different *Sida* species. Protocols proposed in Kress and Erickson, 2007; Theodoris et al, 2012 were used as the starting Protocol for PCR. PCR for each sample was performed in 25 reaction mixture. It contained PCR grade water. PCR buffer, dNTPs, MgC1, forward primers, reverse primers, template DNA, and Taq polymerase (Gotaq, Promega, USA) which were necessary for PCR reaction and concentration of the components and volumes are given in Table 2.2.The master mixture was prepared to reduce the personal errors as the volumes used for each tube is small. It was prepared by adding enough reagents for all number of samples into an Eppendorf and aliquot into separate PCR tubes. DNA in different concentrations was diluted using PCR water to get 20-40 ng. DNA template (1 µl) was added for each PCR tubes. To the negative control l µl of PCR, water was added instead of template DNA. Amplification was carried out on a master cycler thermocycler (Eppendorf) under the temperature profile. Temperature profiles were obtained from the profiles suggested by Theodoris et al, 2012. However, the adjustments were done to the primer annealing temperatures using calculated annealing temperature due to unsuccessful initial amplification. The products obtained were loaded into a 1% (w/v) gel containing ethidium bromide and run in 50 V for visualization of amplified bands.

2.8 Estimation of fragment size of PCR products

Agarose gel (1% (w/v)) electrophoresis was carried out with 5ul of diluted 100 bp molecular weight ladder marker and 5 ul of mat K, psbA-trnH, and PCR products separately. Mixed with 1 µl of loading dye were loaded into the wells. The samples were electrophoresis at a constant voltage of 45 V in 1 X TAE buffer for 2 hours. The separated bands were visualized under UV transilluminator (312 nm wavelength). Fragment sizes were estimated relative to the ladder marker.

2.9 Purification of DNA from gel

2.9.1 Low melting agarose gel electrophoresis

The low melting agarose gel matrix (1% w/v) was prepared by adding low melting agarose powder to the TAE (1 X) buffer (see Appendix A) and melting in the microwave oven. The mixture was slightly cooled. Ethidium bromide (2 µl) was added into the agarose solution. Gel casting tray was placed on a leveled surface and the comb was placed 0.5 mm above the tray. The solution was poured into the gel tray without trapping air bubbles. The solidified gel was transferred into an l X TAE buffer in the gel electrophoresis tank. The PCR products were run at a constant voltage of 45 V for two hours.

2.9.2 Purification of PCR products

The required DNA band was excised from low melting agarose gel by preventing, contamination with other DNA bands using a scalpel under a UV transilluminator. The gel slice was transferred into a pre-weighed DNase free Eppendorf tube. Agarose gel plus Eppendorf tube was weighed. The weight of the gel slice was calculated. DNA, purification from agarose gel was carried out using Illustra GFX PCR and Gel Band Purification kit (GE Healthcare UK Limited, Buckinghamshire). Capture buffer type 3 was added into the Eppendorf tube containing a gel slice. The volume of capture buffer type 3 was 1 µl for each 1 mg of the gel slice. The mixture was incubated at 60ºC by frequent inversion until the gel slice dissolves completely. The color of the solution appeared in yellow. Buffer 3 (600 µl) with a sample mix was transferred to the assembled GFX column and collection tube and incubated at room temperature for 60 seconds. The Micro spin column

plus collection tube containing the sample mix was spin at 14000 rpm for 30 seconds. The flow-through was discarded and the binding step was repeated until the entire sample was loaded. Then 500 µl of wash buffer type 3 was added. The Micro spin column with a collection tube was spin at 14000 rpm for 30 seconds. The collection tube was discarded and the Micro spin column was placed inside a DNase free 1.5 ml Eppendorf tube. Elution buffer type 6 (20 µl) was added into the Micro spin. Column placed inside a clean DNase free tube and incubated at room temperature for 60 seconds. Then it was spun at 14000 rpm for 60 seconds. The purified DNA sample was observed in 1% agarose gel and quantified using BioSpec-nano spectrophotometer and stored at -20 °C until use.

2.10 DNA sequencing

2.10.1 Cycle sequencing

Cycle sequencing for the purified gel band was carried out by using the Big Dye Terminator v 3.1 cycle sequencing kit according to the manufacturer's instruction (Thermo Fisher). The terminator ready reaction mix contained DNA pol thermosequenase, dNTP, dd NTP, reaction buffer. Forward and reverse primers of rbcLa, psbA3`f forward, trnHF reverse, ITS-2 forward, ITS-2 reverse, rnatk forward, and matK reverse primers were used for sequencing, The sequencing reaction mixture preparation is shown Table 2.7.

Table 2.7: Components of the sequencing reaction mixture and their dilutions

Reagent	Sequencing Reaction dilution
	Eight reactions (µl)
Ready reaction mix (2.5x)	1.00
Dilution buffer (5x)	1.50
Template (200ng/µl)	1.00
Primer (0.8 ng/ µl)	2.00
PCR water	4.50
Total	10.00

The sequencing reaction was carried out in Veriti™ thermocycler (Applied Biosystems) according to the temperature profile which is suggested by Kress et al, with modifications as shown in Table 2.8.

Table 2.8: Temperature profile for the cycle sequencing program

Step	**Temperature**	**Duration**
Initial denturation	95 ºC	15 seconds
Denaturation	94 ºC	15 seconds
Primer annealing: psbAf-trnHr	50 ºC	15 seconds
matKf,matKr,rbcl f,rbcl r	52 ºC	15 seconds
DNA strand extension	60 ºC	4 minutes
Final hold	15 ºC	Until tubes remove from the thermocycler
Number of Cycle:30		

2.10.2 Post reaction cleanup

PCR products were transferred into 1.5 ml Eppendorf tubes. Master mix 1 was prepared from 10 µl of PCR water and 2 µl of 125 mM EDTA (appendix 1) per reaction. Master mix (12 µl) 1 was added to each reaction containing 10 µl of reaction. The contents were mixed. The master mix 2 was prepared from 2µl of 3M NaOAc pH 4.6 (appendix A) and 50 µl of absolute ethanol per reaction. Master mix 2 (52 µl) was added to each reaction. The content was mixed and incubated at room temperature for 15 minutes. It was spun at 12 000g for 20 minutes at room temperature. The supernatant was decanted. Then 250 µ1 of 70% ethanol was added and spun at 12 000g for 10 minutes at room temperature. The supernatant was decanted. Finally 14 µl of ABI/ Hi-Di formamide was added. Samples were mixed and transferred to 96-well- plates. The 96-well- a plate was denatured at 95 °C for 5 minutes in a thermocycler. Snap chilled and used for electrophoresis in ABI 3500Dx sequencer.

2.10.3 Sequencing in ABI 3500 Genetic Analyzer

After the Cycle sequencing reaction, the automated sequence detection was carried out for each amplified product of forward and reverse primers. A long run sequencing program was run in ABI 3500 Genetic Analyzer sequencer following manufacturer's instructions (Applied Biosystems, Foster City, USA). Sequencing was repeated two times with the same primers.

2.11 Sequence analysis and submission to GenBank

The sequence obtained by BigDye terminator was analyzed using Bio Edit (version 7.1.3.0) sequence analysis software, Gen bank database, and NCBI BLAST. Recovered trace files for each of the three markers were imported into Bio Edit and exported individually in FASTA format. Sequences of each region were pre-aligned with the ClustalW multiple alignment tool of the Bioedit software and ambiguously aligned regions were manually corrected. Both sides of' the sequences were trimmed, assembled into contigs, and ambiguities were resolved. All sequences have been deposited in the GenBank database.

2.12 Data analysis

The intra and interspecies pairwise distance defined as the distance within the mean group and distance between the mean group of each species respectively. These were calculated using MEGA 7. The barcoding gaps were calculated as the difference between intra and interspecies divergence. The phylogenetic tree (neighbor-joining method) was constructed using MEGA 7 (Vassou et al 2015).

III. RESULT

3.1 DNA Sequencing and DNA Barcoding

After DNA extraction, PCR amplification and DNA sequencing the sequences of the psbA-trnH, rbcL and matK of forward and reverse primers were obtained as electropherogram files in and format. The quality of the sequences obtained by reverse primers was better than obtained with forwarding primers. After the analysis of sequence using Bio Edit software, ambiguities were identified and manually corrected. Results from three, repeated sequencing reactions were used for this purpose. Ends of the sequences were trimmed. The obtained sequence of barcoding regions of given *Sida* species was analyzed with the Genbank nucleotide database using the blastn program and the generated result was tabulated. In addition to Genbank, BOLD (http://www.boldsystems.org/index.php) database was used to analyze rbcl and matK barcoding sequences of the given *Sida* species. The BOLD (**B**arcode **O**f Life **D**ata System V4) contain plant nucleotide sequence only for rbcl and matK, The generated result from BOLD. The matK result of the Genbank database did not reveal similar *Sida* species. Thus BOLD database was used to identify those species.The nucleotide sequence data of *Sida alanifolia* rbcl and *Sida cordata* psbA-trnH were not available in the Gen bank pubic database. Both Gen bank and BOLD databases could not use to identify *Sida cordata* matK and *Sida cordata* psbA-trnH nucleotide sequences. The rbcl, matK, ITS-2, and psbA-trnH sequences of *Sida* species obtained from the study were made available in the Gen bank database and the accession numbers (Table 3.5) were obtained.

Table 3.5: Accession numbers of deposited sequences

Accession no	Name of the submitted sequence
MG596787	*Sida cordata* voucher MS-63-004-02 5.8S ribosomal RNA gene and internal transcribed spacer 2, partial sequence
MG596788	*Hibiscus micranthus* voucher MS-63-007-01 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence
MG596789	*Sida acuta* voucher MS-63-006-01 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence.
MG596790	*Sida alnifolia* voucher MS-63-002-01 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence.
MG601501	*Sida acuta* chloroplast DNA containing psbA-trnH IGS, specimen voucher MS-63-006-01
MG601500	*Sida acuta* voucher MS-63-006-01 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, partial cds;chloroplast.
MG601501	*Sida acuta* chloroplast DNA containing psbA-trnH IGS, specimen voucher MS-63-006-01
MG601502	*Sida cordata* voucher MS-63-004-02 PsbA gene, partial cds; psbA-trnH intergenic spacer, complete sequence; and tRNA-His (trnH) gene, partial sequence; chloroplast.
MG601503	*Sida alnifolia* voucher MS-63-002-01 ribulose 1-5 bisphosphate carboxylase oxygenase (rbcL) gene, partial cds; chloroplast.
MG601504	*Sida alnifolia* voucher MS-63-002-01 PsbA gene, partial cds; psbA-trnH intergenic spacer, complete sequence; and tRNA-His (trnH) gene, partial sequence; chloroplast.
MG601505	*Hibiscus micranthus* voucher MS-63-007-01 PsbA gene, partial cds; psbA-trnH intergenic spacer, complete sequence; and tRNA-His (trnH)

3.2 Data analysis

Phylogenetic trees were constructed using MEGA 7 for each DNA barcoding markers with sequences obtained for each marker (appendix 2). The generated phylogenetic trees were shown below.

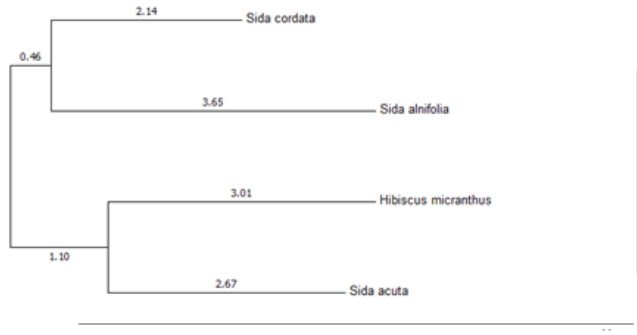

Figure 3.10: The NJ tree constructed from the ITS-2 sequence of given *Sida* species based on Tamura 3- parameter (T92) model (Tamura et al, 2013).

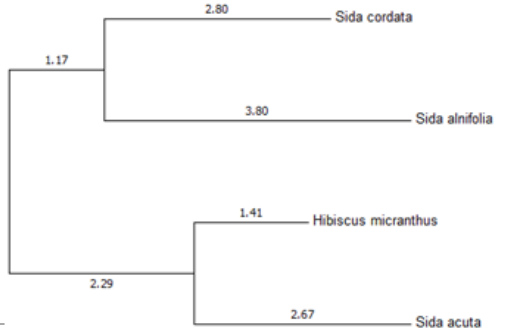

Figure 3.11: The NJ tree constructed from the rbcl sequence of given *Sida* species based on Tamura 3- parameter (T92)

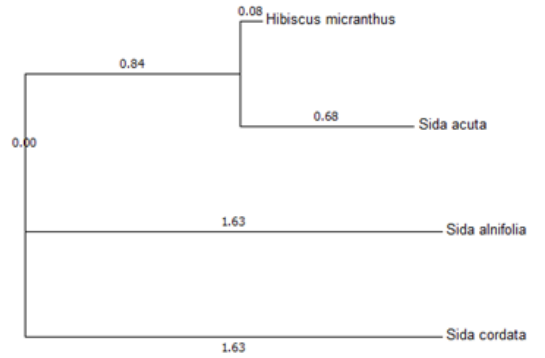

Figure 3.12: The NJ tree constructed from the psbA-tmH sequence of given *Sida* species based on Tamura 3- parameter (T92) model (Tamura et al, 2013).

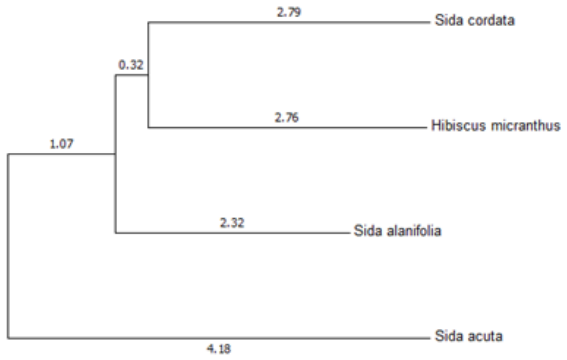

Figure 3.13: The NJ tree constructed from the matK sequence of given *Sida* species based on Tamura 3- parameter (T92) model (Tamura et al, 2013).

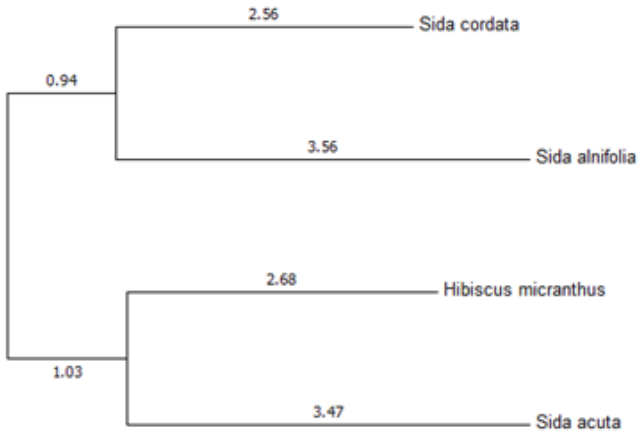

Figure 3.15: The NJ tree constructed from the ITS-2 + rbcl sequence of given *Sida* species based on Tamura 3- parameter (T92) model (Tamura et al, 2013), Evolutionary relationships of taxa

The intraspecies divergence between *Sida* species (table 3.6) was calculated by MEGA 7, using the data generated by the phylogenetic trees shown in figure 3.10-3.13.

Table 3.7: Intra-species divergence between the species of *Sida*.

DNA marker(s)	Intraspecific species divergence
ITS-2	9.97
rbcl	9.37
psbA-tmH	8.67
matK	7.78

IV. DISCUSSION

The constructed phylogenetic trees provide branch lengths of each species, which enable us to analyze the relationship between species and to identify species resolution. Among the DNA bar cording markers, matK unable to resolve, the species efficiently. For an instant, it has out grouped *Sida acuta* from others of the same genus and included *Hibiscus micranthus* as a sister taxa to *Sida cordata*. Since *Hibiscus micrathus* is from different genus it has to be out grouped whereas *Sida acuta* should be branched together with other species of same genus *Sida*. Thus according to the phylogenetic tree, it can be assumed that matK has the least species resolution capacity compared to other DNA barcoding markers used in the study. The other three barcoding marker have well resolved the species, out grouping *Hibiscus micrathus* from genus *Sida* species. However, these phylogenetic trees indicated a closer relationship between *Hibiscus micranthus* and *Sida acuta*.Even though the main objective of the study was to barcode *Sida* species, the *Hibiscus micranthus* was also studied, because it was reported as the most common adulterate to *Sida* species in Sri Lanka. Kumar et al; (2010) showed that due to its similarity in the chemical composition the adulteration of *Hibiscus micrathus* to *Sida* is used in the cases where the *Sida* plants are completely not available. Due to such adulteration, some beneficial quality of *Sida* is being lost. Thus it is important to identify and discriminate *Hibiscus micranthus* from *Sida*. Such a task became failed using morphological identity due to the dried and crushed nature of the samples. Thus DNA barcoding plays a vital role in understanding the relationship between species by analyzing its DNA level. Concerning this; the constructed phylogenetic tree helps to discriminate *Hibiscus micranthus* from other *Sida* species. The *Hibiscus micranthus* was resolved by the barcoding markers. Three out of four barcoding markers showed a closer relationship between *Hibiscus micranthus* and *Sida acuta*, which may share a common region of nucleotide sequences. Using the obtained branch length (provided by the phylogenetic tree) the distance between *Hibiscus micrathus* and *Sida acuta* was calculated as shown below in table 3.6.

Among the used barcoding markers ITS-2 obtained the highest distance between species *Hibiscus micrathhus* and *Sida acuta*, followed by rbcl. This showed that ITS-2 has a higher capacity to distinguish these two species compared to other barcoding markers and can be used as a foremost DNA barcoding marker to overcome adulteration of *Hibiscus micranthus* with *Sida acuta* and other species of genus *Sida*. The distance between species cannot be calculated in MEGA 7 for a single marker at a time, thus it was calculated manually. However using MEGA 7, species resolution of a single barcoding marker can be calculated concerning sequences that are specific to other barcoding markers. Such a way of computed distance is known as intraspecific distance. It is defined as the mean pairwise distance within groups of taxa. The computed data enable to compare the efficiency of the barcoding marker. Vassou et al., (2015) showed that DNA markers with the highest intraspecific sequences divergence can be determined as the 'best' barcoding marker.Out of the used barcoding markers, ITS-2 obtained the highest intraspecific sequences divergence, followed by rbcl. Therefore to obtain more species resolution the combination of ITS-2+rbcl barcoding marker (with highest species resolution) was used to construct a phylogenetic tree.

The distance between species *Hibiscus micranthus* and *Sida acuta* (2.68+3.47) is 6.15. Barcording markers ITS-2 and rbcl showed better species resolution than other markers. These two markers were complementary to each other in species resolution. This complementary nature was reflected in the phylogenetic tree that was constructed for the ITS-2 +rbcL two marker combination. Therefore, ITS-2+ rbcL marker combination is suggested for the DNA barcoding of the samples in the future that are suspect to either adulterate or substitutefor *Sida* species.Concerning the morphological identification, although reproductive characters of a plant are indispensable for species identification, due to the rare flowering nature of *Sida* species, flowers or fruits were not found during the study. Live specimens of *Sida* were identified using the vegetative morphological characteristics as given in the literature to some extent. Although there are several minor differences, most of the vegetative characters of *Sida alanifolia* showed many similarities to a species in the same genus *Sida acuta*. During specimen voucher preparation, the morphological characters of the given *Sida*

species were observed. However, the researcher failed to differentiate these species by such morphological characters. Morphological identification of dry specimens based on the comparison of herbarium specimens also exhibited similar characters in both species. That might be the major reason to use *Sida acuta* and *Hibiscus micrathus* for adulteration of *Sida alanifolia* (Kane et al., 2012). Thus to overcome such adulteration further identification using DNA barcoding was carried out.

While extracting the plant genomic DNA Even though a considerable yield of DNA with the DNA extraction method (Chen and Ronald, 1999), PCR amplification was not successful. Smeared DNA observed in agarose gel for the genomic DNA extracted using such a method (figure 4.1). While investigating the reason behind such smeared DNA, the author came up with different options such as;

a) DNA damage during the extraction process
b) DNA degradation while storing
c) Inadequate / over incubation time of samples with different chemicals of the extraction process.
d) Nature of the given plant sample

Modification of the experimental procedure in PCR using different annealing temperatures, the concentration of MgCl2 and dNTP failed to amplify the barcoding regions of DNA. When the extracted DNA was introduced to properly working another PCR reaction, amplification inhibition was observed by reduced yield. For further clarification, DNA was extracted from a live *Sida* plant leaves as a control next to the given plant sample, using the same quantity of the samples. The extraction process was carried out keeping all other parameters constant. Crushing with nitrogen was reduced and no vigorous vortex was carried out during the process and stored in TE buffer. A non smeared bright band for the control process revealed that the smeared DNA is due to the nature of the sample. Next, PCR was carried out for the smeared DNA, using the live leave extracted DNA as control. Successful PCR amplification was observed in control DNA. These results confirmed that a contaminant in extracted DNA has inhibited the PCR reaction.Although the extraction buffer has added immediately after crushing; the accumulation of phenolic compounds had not eliminated. Therefore, the DNA extraction protocol was optimized by applying several measures to stop the accumulation of polyphenol and proteins to obtain DNA with minimal amounts of PCR inhibitors. To remove polyphenol compounds of the leaves, instead of 2% w/v PVP-40, the percentage was increased to 3%. The protocol suggested by Chen and Ronald, 1999 removes proteins only using chloroform: isoamyl in 24:1 ratio. For further protein precipitation, phenol was introduced to CHISAM, making phenol: chloroform in a 1:1 ratio. Initially chloroform: isoamyl was taken as 24:1 and the mixture: phenol ratio is adjusted to 1:1 in the final solution The CHISAM treatment was carried out once in the Chen and Roald (1999) protocol. However, CHISAM treatment was also carried out twice in another sample of *Sida* DNA extraction to maximize the removal of protein.

When phenol is added and increase the number of CHISAM treatment bands of DNA are faint. Accordingly, this shows that the addition of CHISAM has considerably reduced the DNA yield, but the quality has improved. Concerning the study by Kazi et al., (2013) on the impact of different sample methods on DNA quality indicated air drying and alcohol pre-treatment can result in significantly lower DNA yields, and lower PCR success as well as more fragmented DNA. The study suggested that natural drying over longer periods in lower heat can eliminate such damages to DNA and allow it to undergo follow up experiments such as extraction, PCR as well as sequencing.Under such processing and manufacturing conditions, DNA can become degraded and fragmented, resulting in broken strands of DNA. If one of these breakages occurs in between the annealing positions of PCR primers then an amplicon will not be formed and the reaction will fail (Alaeddini, Walsh and Abbas, 2010). The likelihood of breakage between primer annealing positions is positively correlated to the length of the amplicon. This result shows the problem of using longer regions, such as DNA barcode regions, for the identification of processed plant material (Coyne et al., 2006).Also while drying many primary and secondary metabolites found in plants, polyphenols and polysaccharides affect DNA extraction most severely through DNA oxidation by covalently binding to nucleotides and by inhibiting enzymatic reactions during PCR. Also, particular leaf types and textures can similarly hinder DNA extraction (Kress and Erickson, 2007). Even though many studies showed PCR failure occurs in DNA samples that undergo drying process, in contrast, Pusch, (1998) showed that nuclear and commonly used plastid markers are available for PCR amplification in the same ratio as in fresh tissue even after their preservation. In choosing the better DNA extraction protocol, it is also important to consider acquiring DNA with high purity despite high yield, because failure to clean DNA of polyphenols and polysaccharides can result in negative PCR results (Bacich et al., 2011). This can occur due to the PCR inhibiting properties of chemicals even in non-degraded DNA samples.

Best performing DNA extraction protocols for dried DNA are those that combine high purity with high DNA yield, such as a combination method of CTAB with silica binding. Sarkinen et al., (2012) exhibit these methods as much promise, and focus should be given to their further development and upscaling. After the failure of PCR after several attempts of modification in the DNA extraction protocol and optimization of PCR protocol, genomic DNA extraction using a kit was adopted from Sarkinen et al., (2012). Ceygen plant genomic extraction kit that incorporated enhanced CTAB method with silica binding was used to extract the genomic

DNA of given *Sida* species and PCR was carried out. Even though sheared bands were obtained for extracted genomic DNA the PCR was successful. It is concluded that the genomic DNA of *Sida* was damaged during the drying process thus alteration in the extraction process cannot yield a nonfragmented (non smeared) DNA. However, a more modified and enhanced DNA extraction protocol can increase the quality of the DNA which can be used for post-extraction experiments like PCR and sequencing. Additionally, the concentration of DNA was determined using two methods in the study. Those are measuring the absorbance of DNA in a UV-visible spectrophotometer and using the BioSpec-nano spectrophotometer. Quality evaluation of DNA can be done using agarose gel electrophoresis in addition to these two techniques. To roughly estimate DNA concentration, UV-visible spectrophotometry is useful, but its accuracy is low for low DNA yields. For Cycle sequencing DNA concentration should be accurate (Costion et al., 2016). After GFX purification PCR product yield seemed less in agarose gel electrophoresis. Also, the quality of DNA is paramount for DNA sequencing. Thus, purified PCR product quantification and quality evaluation were carried out using both agarose gel electrophoresis and BioSpec-nano spectrophotometry.

Unlike fresh samples, isolation of genomic DNA from dried plant parts and powders that are suitable for PCR can be challenging due to the degradation of DNA and the presence of unknown PCR inhibitors (Lee et al., 2016). Though the quality of DNA as determined from the OD260/280 ratio and agarose gel electrophoresis was variable, PCR amplification was successful with all the market samples after the standardization of DNA quantity. In general re-amplification of failed samples by using the lowest amount of template DNA was successful, probably due to the dilution of PCR inhibitors.While discussing the polymerase chain reaction of the study, according to the results the rbcL, matK, ITS-2 and psbA-trnH primers are suitable to amplify rbcL, matK, ITS-2 and psbA-trnH regions of the chloroplast DNA of *Sida* species. Amplification of the psbA-trnH, matK, ITS-2, and rbcL was carried out with designed primers. PCR for the first extracted DNA sample was not successful due to the low DNA quality. In this study PCR amplification success rates were different for the four standard DNA barcodes, rbcL(100%) >ITS-2(99.4%) > psbA-trnH (97.2%) >*matK* (59.8%) in four samples, *matK* was the most difficult one to be amplified. The amplification success of *matK* was approximately 60% in angiosperms with 'universal' primer pair (3F/1R), which is used in this study. In contrast, Ashfaq et al., (2013) in their study showed that the use of another set of primers (390F/1326R) could increase amplification and sequencing success by approximately another 10%.

To amplify the sequence of matK, they used KOD FX DNA polymerase (more efficient at amplifying the sequence and produce high-quality fingerprints) and 3000ng template DNA, which was much higher than the normal amount of template DNA (Tamura et al., 2011). Initial PCR produced a low yield of the amplified product. Therefore, minor adjustments were carried out for annealing temperature based on the base composition of the primers. Moreover, MgCl2 and dNTP concentrations were increased. All of these adjustments resulted in intense bands. As primer dimers or nonspecific bands were not observed in agarose electrophoresis further optimization was not required.DNA barcoding mainly depends on the success rate of PCR amplification and DNA sequencing. Although the plant samples contained polysaccharides and other secondary metabolites, PCR amplification and bidirectional sequencing of rbcL, matK, psbA-trnH, and ITS2 barcode markers were successful with all the 16 accessions collected.The expected band size of the PCR products of psbA-trnH, rbcL, ITS-2, and matk was 300-800 bp, 600-800 bp, 300-400 bp, and 800-1500 bp respectively. According to Figure 4.8, the PCR products obtained were on expected sizes. As predicted, there was no size variation in rbcL marker among the given sida species. However, there was a different range of length variations in matK, psbA-trnH, and ITS2 markers. BLAST search in Genbank and BOLD Systems databases recovered the expected *Sida* species.Concerning the PCR product purification, even though the PCR product itself could be purified using the GFX column; gel band purification was carried out as high purity of the product may have a positive impact on the sequencing reaction. This technique removes excess primers and dimmers if they are present in the PCR products. According to figure 3.9 band intensity of purified DNA is low. DNA purification further confirms the reduction of DNA amount as listed in table 3.2. Therefore GFX purification has reduced the DNA yield.

Also, all the primers resulted in sequences in good quality and the open reading frame was found to be intact for the coding markers. However, three reaction sequences had low signal strength, dye blobs, and low fragment length. Poor template quality is the most common cause of sequencing problems for instance noisy data (peaks under the peak), no usable sequence data, and weak signal. Potential contaminants include; proteins, RNA, chromosomal DNA, excess PCR primers, dNTPs, enzyme, and buffer components (from a PCR amplification used to generate the sequencing template) and residual salts (Ojeda et al., 2014). But in these reactions, at least one primer has given a good quality sequence. The problem might have occurred during the post-reaction cleanup. The obtained sequences have no mixed peaks indicating that all the given samples are single species. Thus this sample can be DNA barcoded without amplification of species-specific markers or cloning of PCR products.

Both Genbank and BOLD databases could not able to identify *Sida cordata* matK and *Sida cordata* psbA-trnH. The study by Vassou et al, (2015) indicated that the matK barcoding marker cannot differentiate *S.cordata* species. Moreover, there is no nucleotide database available in Genbank for *S.cordata* psbA-trnH. Since BOLD have nucleotide sequences only for rbcl and matK markers, there was no source of nucleotide database to identify the psbA-trnH region of *Sida cordata*. The result of data analysis revealed that ITS-2 marker was able to differentiate all the given *Sida* species, with the highest distance between the species and intraspecies divergence. The intraspecies divergence was calculated to identify the better combination of the marker. The intraspecies divergence was highest to the marker combination of ITS-2 + rbcL. Both individual marker data, as well as two marker combination data, were used to construct phylogenic trees (Srirama et al., 2010). Among the constructed phylogenetic trees, the ITS-2 and rbcL generated trees showed better species resolution. Hence the combination of ITS-2 + rbcl was determined as a better marker combination for the DNA barcoding of *Sida* species of Sri Lanka.

In split, the DNA barcording method depends mainly on differentiating intra-specific from inter-specific genetic variation. However, the ranges of such variation may differ between taxa. Thus it is difficult to resolve new species or recently diverged species that have arisen through hybridization (Santhosh Kumar et al., 2015). There is no universal gene of DNA barcoding that is conserved in all species of life. Thus DNA barcoding should be done using an already established reference sequence of validated specimens. This is comparatively a complex process involving a diverse group of scientists and institutions. The fundamental limitation of barcoding is the short (500-1000bp) sequence used in the methodology, which limits their resolution power in discriminating deep branches in phylogenies (Bharati, 2016). Even though the DNA barcoding technique is widely used today for species discrimination still many controversies exist over the validation of DNA barcoding. First, rather than enhancing the traditional morphology method, DNA barcoding tends to diminish (Verdcourt, 2004). Techen et al., (2014) indicated solely depending on genetic divergence for species determination could result in incorrect species recognition. However, it must be relied on in the open possibility that these sequences help to increase the taxonomic coverage, which can be on exceptional resource for taxonomy, systematic biology, and maybe equally useful (Poovitha et al., 2016).

In conclusion, the findings of this study indicate that currently used primers of psbA-trnH, ITS-2, rbcL, and matK in DNA barcoding, can amplify the regions of *Sida* DNA. DNA sequencing was successful for all four regions. Blastn search aligned *Sida* species with a plant in the same genus, whereas matK region was not accurately aligned. The results highlight that among the four regions used in this study, ITS-2 and rbcL barcode regions could serve as the best DNA barcodes for the *Sida* species of Sri Lanka. The regions for these species were made available at the Genbank database. These DNA barcodes will contribute to the identification, conservation, and quality control of valuable Sri Lankan medicinal plant species of *Sida*.As the advantages and limitations of DNA barcoding become evident, it is clear that taxonomic processes incorporating DNA sequencing will achieve maximum efficiency at species identification. DNA barcoding enables us to speed up the work of taxonomists in species identification. Despite some drawbacks of using DNA barcoding, the study reported here showed the success of using the barcoding region in distinguishing *Sida* species of Sri Lanka and resolved species adulteration of Babila in the natural drug market. The study recommends that in the future, efforts should be taken to develop nuclear barcodes to complement the barcoding regions that are currently in use for even more accurate barcoding results.

REFERENCE

[1]. Alaeddini R, Walsh S. and Abbas A. Forensic implications of genetic analyses from degraded DNA-A review. Forensic Science International: Genetics. 2010; 4:148-157.

[2]. Ashfaq M, Asif M, Anjum Z. and Zafar Y. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (Gossypium: Malvaceae). Molecular Ecology Resources. 2013; 13:573-582.

[3]. Bacich D, Sobek K, Cummings J, Atwood A. and O'Keefe D. False negative results from using common PCR reagents. BMC Research Notes.2014; 4:457-467.

[4]. Baracho S, George Y, and Maria A. A New Record Of Sida Glutinosa (Malvaceae), A Rare Species Of The Caatinga In Northeastern Brazil, With Lectotypifications And Taxonomic Notes On The Allied Sida Glabra. Phytotaxa 2016: 37.

[5]. Bharati, K.. Identification of Indian Sida through mericarp. Pharmacognosy Journal. 2016;8:490-496.

[6]. Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y. and Leon C. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLoS ONE. 2010; 5: 614-618.

[7]. Cheung W, Hubert N. and Landry B. A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. Genome Research. 1993; 3:69-70.

[8]. Costion C, Lowe A, Rossetto M, Kooyman R, Breed M, Ford A. and Crayn D. Building a Plant DNA Barcode Reference Library for a Diverse Tropical Flora: An Example from Queensland, Australia. Diversity. 2016; 8:5-10.

[9]. Coyne K, Hare C, Popels L, Hutchins D. and Cary S. Distribution of Pfiesteria piscicida cyst populations in sediments of the Delaware Inland Bays, USA. Harmful Algae. 2006;5:363-373.

[10]. Dubey S, Tripathi A, Upadhyay B. and Kumar A. Development of conventional and real time PCR assay for detection and quantification of Rhizoctonia solani infecting pulse crops. Biologia. 2016; 71:89-95.

[11]. Fazekas A, Burgess K, Kesanakurti P, Graham S, Newmaster S, Husband B, Percy D, Hajibabaei M. and Barrett S. Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE. 2008; 3:178-184.

[12]. Gunatilaka, A., Arseculeratne, S., and Panabokke, R. Studies on medicinal plants of sri lanka. part 14: toxicity of some traditional medicinal herbs. Journal of Ethnopharmacology, 1985; 13:323-335.
[13]. Hebert P, Cywinska A, Ball S. and deWaard J. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences. 2003; 270:313-321.
[14]. Hilu K. and Liang H. The matK Gene: Sequence Variation and Application in Plant Systematics. American Journal of Botany. 1997; 84:830-836.
[15]. Jian. "A Successful Case Of DNA Barcoding Used In An International Trade Dispute". DNA Barcodes 2.1. 2014.
[16]. Kane N, Sveinsson S, Dempewolf H, Yang J, Zhang D, Engels J. and Cronk Q. Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. American Journal of Botany. 2012; 99:320-329.
[17]. Kazi T, Hussain N., Bremner P., Slater A. and Howard C. The application of a DNA-based identification technique to over-the-counter herbal medicines. Fitoterapia. 2013; 87:27-30.
[18]. Koolman J. and Roehm K. Color atlas of biochemistry. 1st ed. Stuttgart: Thieme. 2013.
[19]. Kress W. and Erickson D. A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS ONE. 2007; 2:508-511.
[20]. Kumar D. and R. Kundapur R.Biomedical Applications of Natural Proteins. 1st ed. New Delhi: Springer India.2015.
[21]. Kumar, K., Ramachandra Setty, S. and Narsu, L. Pharmacognostic and Phytochemical Investigations of Stems of Hibiscus micranthus Linn. Pharmacognosy Journal, 2010; 2:21-30.
[22]. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit, S, Barraclough, T. and Savolainen, V. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences, 2008;. 105:2923-2928.
[23]. Lee S, Mahat M, Nazre M. and Mohamed, R. DNA Barcoding of the Endangered Aquilaria (Thymelaeaceae) and Its Application in Species Authentication of Agarwood Products Traded in the Market. PLOS ONE. 2016;11:154-171.
[24]. Lee. DNA Barcoding Of The Endangered Aquilaria (Thymelaeaceae) And Its Application In Species Authentication Of Agarwood Products Traded In The Market. PLOS ONE.2016;87-90.
[25]. Mishra P, Kumar A, Nagireddy A, Mani D, Shukla A, Tiwari R. and Sundaresan V. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnology Journal.2015; 14: 8-21.
[26]. Montelone B, Prakash S. and Prakash L. Recombination and mutagenesis in rad6 mutants of Saccharomyces cerevisiae: Evidence for multiple functions of the RAD6 gene. MGG Molecular & General Genetics. 1981; 184:410-415.
[27]. Newmaster S, Grguric M, Shanmughanandhan D, Ramalingam S. and Ragupathy S. DNA barcoding detects contamination and substitution in North American herbal products. BMC Medicine. 2013; 11:78-82.
[28]. Nithaniyal S, Vassou S, Poovitha S, Raju B, Parani M. and Cristescu M. Identification of species adulteration in traded medicinal plant raw drugs using DNA barcoding. Genome.2017; 60: 139-146.
[29]. Obah, J O Akerele and O Obasuyi. Antimicrobial Activity Of The Ethanol Extract Of The Aerial Parts Of Sida Acuta Burm.F. (Malvaceae). Tropical Journal of Pharmaceutical Research. 2007;7:451-455.
[30]. Ojeda D. I. DNA Barcodes Successfully Identified Macaronesian Lotus (Leguminosae) Species Within Early Diverged Lineages Of Cape Verde And Mainland Africa. AOB PLANTS.2014;47:784-794.
[31]. Palhares R, Drummond M, Brasil, B, Cosenza G, Brandao M. and Oliveira G. Medicinal Plants Recommended by the World Health Organization: DNA Barcode Identification Associated with Chemical Analyses Guarantees Their Quality. PLOS ONE.2015; 10:116-124.
[32]. Parveen I, Gafner S, Techen N, Murch S. and Khan I. DNA Barcoding for the Identification of Botanicals in Herbal Medicine and Dietary Supplements: Strengths and Limitations. Planta Medica.2016; 82:1225-1235.
[33]. Poovitha S, Stalin N, Balaji R, Parani M. and Bank M. Multi-locus DNA barcoding identifies matK as a suitable marker for species identification in Hibiscus L. Genome. 2016; 59:1150-1156.
[34]. Pusch C. Repair of degraded duplex DNA from prehistoric samples using Escherichia coli DNA polymerase I and T4 DNA ligase. Nucleic Acids Research. 1998;26:857-859.

R. Rajphriyadharshini, et. al. "Development of DNA barcodes for Medicinal plant Sida species in Sri Lanka." *International Journal of Pharmaceutical Science Invention*, vol. 09(6), 2020, pp 24-34. Journal DOI- 10.35629/6718