Eco. Env. & Cons. 30 (2) : 2024; pp. (957-964)
Copyright@ EM International
ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2024.v30i02.089

Identification and classification of medicinal plant *Cassia* species through DNA barcode to validate phylogenic analysis of Rbcl, Matk, Ycflb, trnH-psba and ITS marker

Jeevitha S. and Anandan R.*

Department of Genetics & Plant Breeding, Faculty of Agriculture, Annamalai University, Annamalai Nagar 608 002, India

(Received 7 January, 2024; Accepted 23 March, 2024)

ABSTRACT

Cassia species are considered as a valuable source for our Indian medical system. Inter and intra species of *Cassia* is genetically variable and its necessary to identify and discriminate individuals in order to exploit its maximum potential. DNA barcode technique of well-studied universal markers were used to amplify the genomic region of several *Cassia* species. The amplified regions were sequenced and the coding's were blasted and studied for any significant differences. The results from the generated *Rbcl, MatK* and *Ycf1b* sequences were almost identical to *Senna* species that is closely related to *Cassia* species. The phylogenetic tree for five species of *Cassia* was constructed using three markers (*Rbcl, MatK* and *Ycf1b).* Among the markers used *Rbcl*_1 shows the best resolution for *S. auriculata*. *MatK*_3 has showed better species resolution for *S. occidentails* than other markers. The implication of these results can be used in species identification of wild and native genotypes of medicinal plant from various geographical locations.

Key words: *DNA Barcode, Phylogenetic diversity,* Cassia, *Marker, BLAST*

Introduction

The usage of medicinal plants is expanding from the periphery to the mainstream as more people look for treatments and health strategies free from the adverse effects of synthetic drugs. Over 3000 plants are acknowledged as having therapeutic use in India. Over 6000 plants are thought to be used in traditional, folk, and herbal medicine in India (Danish *et al*., 2011). These medicinal herbs are becoming more and more important on a global scale as a result of their increased traditional use and social acceptance. They are also widely recognized and have few side effects. The international trade of medicinal plants and their products was estimated to be USD 60 billion in 2010, and by 2050, it is anticipated to reach USD 5 trillion (Zahra *et al*., 2020). *An essential medicinal herb utilised in many conventional medical systems, including Ayurveda, is Cassia fistula L* (Mwangi *et al*., 2021). *Cassia* sp. belong to the family Fabaceae and grows all over the tropical countries (Jain, 1968). It is an annual shrub and grows in feral lands. The plant's leaves, seeds, and roots are said to have medicinal properties. Several chemical compounds such as Anthraquinone glycosides, Naphthopyrone glycosides, Phenolic compounds, Flavonoids etc.

(Assistant Professor- Biotechnology)

have been isolated from this plant and is useful for treatment of leprosy, ringworm infection, ophthalmic, skin diseases and liver disorders (Singh *et al*., 2013). The seeds of *Cassia* species are commonly used for medicinal purposes whereas, leaves, bark, root, and flower are also used to cure rheumatism, liver disease, diabetes, urinary tract disorders, constipation, jaundice, ulcer, leprosy, skin and conjunctivitis disorders (Hakkim *et al.,* 2007 and Habtemariam, 2013). The inter-specific genetic variation of *Cassia* species are greater than the intra – specific variation, which emphasize the need to identify and recognize every species (Deepak *et al.,* 2018). The poor capacity of humans to recognise and retain morphological variation substantially hinders current approaches to the study of biological diversity. We will require up to 100,000 taxonomists only to maintain our ability to recognise the 10 to 100 million species on Earth once they have all been described, as few taxonomists can consistently identify even 1,000 species. This sobering reality is what inspired DNA barcoding, a novel method of species identification.

The Consortium for the Barcode of Life (CBOL) is an international initiative devoted to developing DNA barcoding as a global standard for the identification of biological species. ,Building the DNA barcode reference libraries, sequencing facilities, informatics platforms, analytical protocols, and international cooperation necessary to inventory and assess biodiversity are the goals of the research alliance known as iBOL (International barcode of life).The vision of CBOL is to increase the rate of conserving the DNA barcodes of known and newly discovered plant and animal species. It also plays a role in establishing a public library of sequences and developing portable devices for DNA barcoding. Since 2008, the molecular identification and classification technique known as DNA barcoding has been used to identify and classify medicinal plants (Yu *et al*., 2021). It is a useful tool for verifying the authenticity of pharmaceutical products and a technique for identifying species using brief DNA sequences that are conserved within species but different between species (Purushothaman *et al*., 2014). Regardless of the plant's physical condition, DNA barcoding could identify it with accuracy, however barcoding faces the difficulty of changing marker loci. (Khan *et al*., 2019). An optimal barcode should be variable enough to discriminate closely related species and easy to handle critical experiment at significantly low cost. The flanking regions of the barcode are conserved enough for precise primer designing which could help for further successful PCR and genome sequencing (Dong *et al.,* 2015). Plant barcoding is used to distinguish between unknown specimens and known species and to share sequences with the community in order to clarify its taxonomy. To detect adulteration in Cassia, four common DNA barcode regions—ITS, matK, rbcL, and psbA-trnH-were employed (Seethapathy *et al*., 2015). A standard portion of DNA is sequenced as part of DNA barcoding using plastid coding matK markers, which is used to identify angiosperms (Prasath *et al*., 2017). Understanding morphological, physiological, molecular, biochemical, discrimination of various *Cassia* species is essential for effective usage of plant species for medicinal purpose. The advancement in the molecular techniques helps to explore more than 400 plus *Cassia* species that are currently being used for medicinal purposes (De Britto *et al.,* 2010). Recent developments in molecular plant identification using DNA sequence data enable accurate identification of plant species from herbal medicines using defined DNA markers (de Boer *et al*., 2015). Genetic sequences obtained from DNA barcoding have also been used to create phylogenetic relationships, which are used in the study of phylogenetic community ecology (Kress *et al.,* 2010). With this background our research to sequence the known chloroplast genomic region of *Cassia* species to compare and explore any difference in the genomic sequences between the species and to elucidate the phylogenetic relationships between the species.

Materials and Methods

Plant Materials

The plant materials used for the present study were collected from ICAR-Directorate of Medicinal and Aromatic Plants Research, Gujarat. The collected seed samples of different cassis species such as *C. auriculata, C. alata, C. occidentalis* and *C. tora* were grown in enriched solid medium containing FYM, sand and soil mixture in the ratio of 1:1:2 and were maintained at the pot culture yard, Department of Genetics and Plant Breeding, Faculty of Agriculture, Annamalai University.

Genomic DNA extraction

Following the Doyle and Doyle (1990) methodology,

the genomic DNA was isolated from the leaf sample using the Cetyl Trimethyl Ammonium Bromide (CTAB) method. The quality of the extracted DNA samples was checked using 0.8% agarose gel electrophoresis. The DNA was quantified using UV Nano Drop (Spectrophotometer, Shimadzu, Japan).

PCR amplification

Extracted DNA samples were amplified using a set of DNA barcode markers such as *Rbcl*, *MatK*, *ITS*, *trnH-psbA* and *Ycf1b* in a thermo cycler (96 well-Applied Biosystem). The marker sequences, annealing temperature and sizes of the PCR products were given in Table 1. The total volume of PCR reaction used was 15 µl. DNA sample 2µl, 1µl of each primer reverse and forward, 7µl of PCR master mix and H_2O 4µl was amplified from total genomic DNA using a thermo cycler (96 well-Applied Biosystem) and the programe temperature is given Table 1. The amplified PCR product size was validated using 1.0% of agarose gel and list of primer used for PCR amplification sequence given Table 2.

Sanger Sequencing

The PCR products were sequenced by using Sanger sequencing methodology, Eurofins, Bangalore, India.

Data analysis

The sequences were analysied using the Chromas (version 2.6) and Molecular Evolutionary Genetics Analysis software (MAGA version 6.0). The sequences were blasted using NCBI website. http://blast.ncbi.nlm.nih.gov/Blast.cgi. Phylogenetic relationship were assessed and the phylogenetic chart were established using similarity, differences between intra and inter species. The trimmed sequences were blasted and the results from the NCBI blast were listed in the Table 3. The result shows that the sequences generated were nearly identical to various *Cassia* species and the number of species that are closely identical were given in Figures 1-7. Although we tried to generate 25 sequences across five different *Cassia* species the quality sequences were obtained only in four species leaving the *Cassia angustifolia* species with three markers but the

Table 1. PCR Reaction Primer Details

S. No.	Primer	Initial & Final Denaturation	Annealing	Extension	Cycle	Product Size
1	Rbcl – FRbcl – R	94 ^{0}C – 2 min 94 ^{0}C – 30 Sec	52 ^{0}C – 30sec	68 ^{0}C – 1 min 68 ^{0}C – 10 min	35	600bp
2	Matk – FMatk – R	94 ^{0}C – 2 min 94 ^{0}C – 30 sec	48 ^{0}C – 30 sec 68 ^{0}C – 10 min	68 ^{0}C – 1 min	35	800bp
3	Ycflb – FYcflb – R	94 ^{0}C – 4 min 94 ^{0}C – 30 sec	52 ^{0}C – 40 sec 72 ^{0}C – 10 min	72 ^{0}C – 1 min	34	450bp
4	trnH-psbA – F trnH-psbA– R	94 ^{0}C – 2 min 94 ^{o}C –30 sec	60 ^{0}C – 30 sec 64 ^{0}C – 10 min	64 ^{0}C – 1 min	35	250bp
5	ITS-18S - F ITS-26S – R	95 ^{0}C – 5 min 95 ^{0}C – 30 sec	50 ^{0}C – 30 sec 72 ^{0}C – 8 min	72 ^{0}C – 90 sec	35	-

Table 2. List of primer used for PCR amplification

S.No	Primer	Sequence
1	Rbcl – F	ATGTCACCACAAACAGAGACTAAAGC
	Rbcl – R	GTAAAATCAAGTCCACCGCG
2	Matk – F	GTCCATGTCGAAATCTTGGTT
	Matk – R	ATGATAATGAGAAAGATTTCTGCC
3	Ycflb – F	TCTCGACGAAAATCAGATTGTTGTGAAT
	Ycflb – R	ATACATGTCAAAGTGATGGAAAA
4	trnH-psbA – F	GTTATGCATGAACGTAATGCTC
	trnH-psbA – R	CGCGCATGGTGGATTCACAATC
5	ITS-18S - F	GAACCTTATCGTTTAGAGGAAGG
	ITS-26S – R	CCGCCAGATTTTCAGGCTGGGC

other two markers have not showed any result for five *Cassia* species.

Results and Discussion

The resulted sequences were quality checked and some of the ambiguous nucleotides on both the ends were trimmed carefully using the chromas software. The trimmed sequences has been uploaded in the NCBI website for the future study purpose and the accession number were listed in the Table 2.

The results from the NCBI for the trimmed sequences were listed in Table 3. The result shows that the sequences generated were similar to various *Cassia* species and those identical were shown in Figure 1-7. While trying to generate 25 sequences with three markers from five different *Cassia* sp. only four species leaving *Cassia angustifolia* generated good quality sequences but the other two markers has not showed any result for five *Cassia* species. *Senna alata* has accession number OP555172(F) and OP555173(R) *Ycf1b_F*primer of *Sennaalata* showed query coverage 94% and the maximum identity is 100%. *Ycf1b_R* primer of *Senna alata* showed query coverage 94% and the maximum identity is 100%.

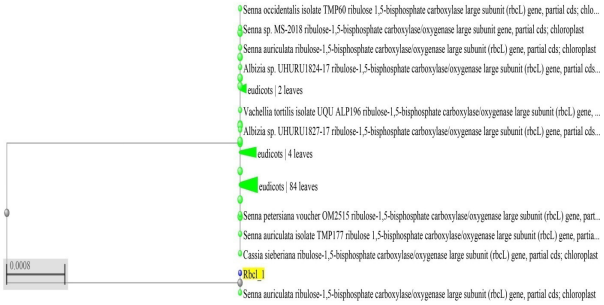

Fig. 1. The phylogenetic tree constructed for the species of *Senna auriculata* (*Rbcl*_1)

Rbcl sequence of *Senna auriculata* showed query coverage 100% and the maximum identity is 100% of accession number OP555107. Maximum identity 99.34% with query coverage 100% and accession no OP555108(F), OP555109(R) of *MatK* sequence in *S. tora* followed by *Ycf1b* sequence as it showed query coverage and maximum identity 100% in the accession number OP555169 for the same species. Accession number given by NCBI Gen Bank OP555110 has showed query coverage 96% and maximum identity 100%, likewise, the maximum identity and query coverage is 100% for OP555170 of *Ycf1b* sequence in *Senna occidentalis*. The results are interpreted in Table 3. The sequences were multiple aligned to see any major variation in the species level, however there was no proper alignment and no significant results could be obtained.

To study the phylogenetic relationship of the *Cassia* sp. dendogram was constructed based on the sequencing result and it showed the close relation between *Senna* species. Five samples of *Cassia* were subjected to DNA barcoding using universally accepted markers. The already known or deposited sequence of markers (*Rbcl, MatK* and *Ycf1b)* was collected from Gen Bank (NCBI) BoLD Data Base to observe the nucleotide variation among them. The result was supported by genetic divergence studies.

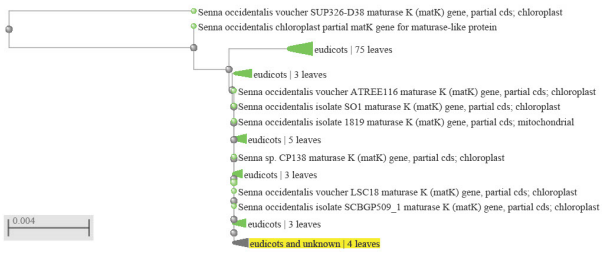

Fig. 2. The phylogenetic tree constructed for the species of *Senna occidentalis* (*MatK_3*)

Table 3. NCBI-BLAST similarity results of five species of *Rbcl, MatK, ITS, Ycf1b and trnH-psbA*sequnces

S. No.	Species	Primer	Maximum Score	Sequence length (bp)	Query coverage (%)	Maximum identity (%)	BLAST "E" value	Accession No.
1	*Senna alata*	Ycf1b_F	651	342	94%	100.00%	1e-176	OP555172
		Ycf1b_R	697	377	100%	100.00%	0.0	OP555173
2	*Senna auriculata*	Rbcl(F&R)	870	471	100%	100.00%	0.0	OP555107
3	*Senna tora*	MatK_F	545	131	100%	99.34%	1e-150	OP555108
		MatK_R	241	130	100%	100.00%	2e-59	OP555109
4	*Senna tora*	Ycf1b(F&R)	1507	816	100%	100.00%	0.0	OP555169
5	*Senna occidentalis*	MatK(F&R)	1192	645	96%	100.00%	0.0	OP555110
6	*Senna occidentalis*	Ycf1b-_F	614	332	100%	100.00%	3e-171	OP555170

The phylogenetic tree for five species of *Cassia* was constructed using three markers (*Rbcl, MatK* and *Ycf1b).* Among the markers used *Rbcl*_1 shows the best resolution for *S. auriculata* (Fig. 1). *MatK*_3 has showed better species resolution for *S. occidentails* than other markers (Fig. 2). *MatK*4_R and *MatK*4_F should resolve *S. tora* but *MatK*_4_R (Fig. 3) shows some similarity for *S. occidentails, MatK*_4_F (Fig. 4) does not match for any five species. Ycf1b_2_F (Fig. 5) exactly found the species *Cassia*alata and Ycf1b_2_R (Fig. 6) does not found any of the five species of *Cassia*. Ycf1b_3_F showed *S. occidentalis* (Fig. 7).

DNA barcoding is an effective method for determining and validating species (Shen *et al*., 2013). A technology called DNA barcoding uses internal species markers made of specified gene regions to identify species. It has gained recognition as an established branch of biodiversity sciences that bridges the conceptual chasm between conventional taxonomy and other branches of molecular systematics. DNA barcoding was originally intended to be a tool for species identification, but it is now used in taxonomy procedures for automated species identification and species demarcation (Hubert and Hanner, 2015). Molecular signatures at conserved barcode region reveal genetic significant and potential for crop improvement (Acharya *et al*., 2022). This study is an attempt to distinguish and discriminate the medicinally valued *Cassia* Species at the molecular level. Species identification in Fabaceae family is highly necessary in order to distinguish invasive species from endangered species like *C. auriculata, C. alata*, *C. occidentalis C. tora*, and *C. angustifolia*. Four DNA barcode regions *ITS*, *matK*, *rbcL*, and *psbA-trnH were used to assess the* the extent of adulteration in the herbal trade of Senna sp. Analysis of market samples revealed considerable adulteration of herbal products: 50 % in the case of *Senna auriculata*, 37 % in *Senna tora*, and 8 % in *Senna alexandrina*

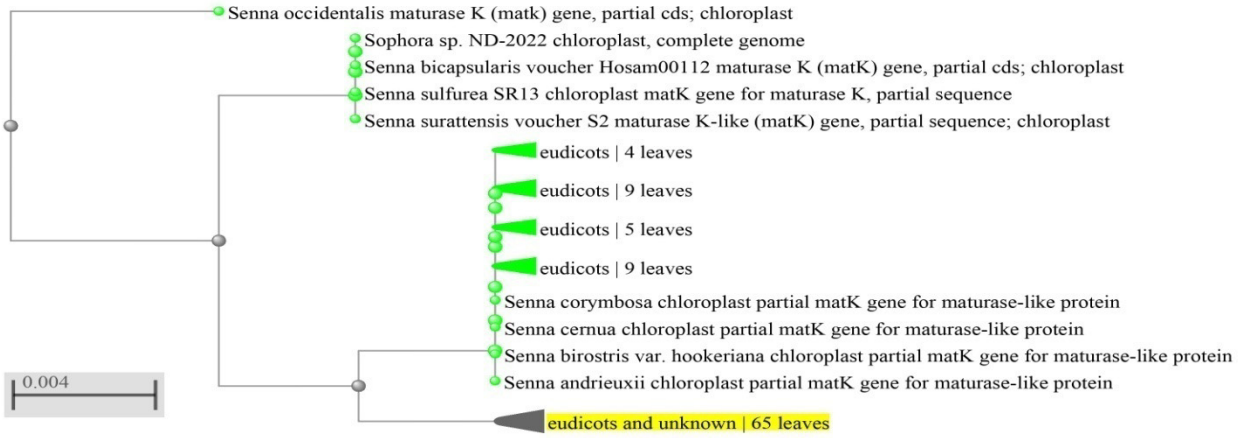

Fig. 3. The phylogenetic tree constructed for the species of *Senna tora* (*MatK_4*_R)

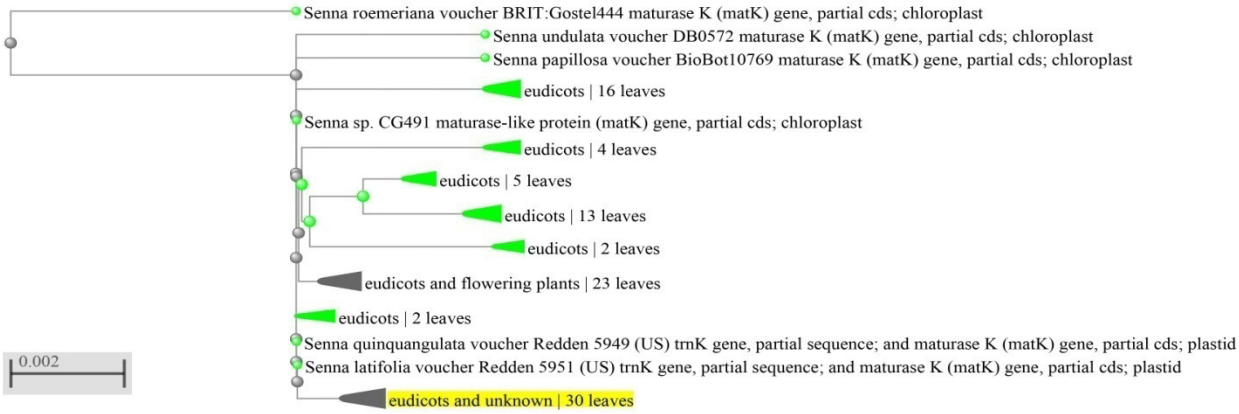

Fig. 4. The phylogenetic tree constructed for the species of *Senna tora* ((*MatK_4*_F)

(Seethapathy *et al*., 2015b). Primers were created to amplify the regions, and their PCR success rates ranged from 82.80% to 98.17%. *Ycf1b* distinguished 357 tree species out of 420, which was slightly better than the combination of *matK* and *rbcl*. In the well-sampled representative plant groups, *ycf1b* outperformed *matK*, *rbcL*, and *trnH-psbA*. It was concluded that the most variable plastid genome region, ycf1a or ycf1b, can serve as a core barcode for land plants (Dong *et al.*, 2015). Molecular identification is the most reliable approach that help researcher for precise characterization of unknown and complex taxon. Universal barcodes *(Rbcl, MatK, ITS, trnH-psbA* and *Ycf1b)* were used for multiple *Cassia* species, the NCBI results identified were perfectly or nearly closed to *Senna* species which shows that the barcodes markers could be used in a lot of native and wild species for genus and species identification. In some cases, the best species discrimination was gained within the same study by using various loci for different plant groups. A study of species from two families growing in the same region, for example, found that *ITS* discriminated better between species from the family Poaceae, while *matK* worked much better with species from the family Chenopodiaceae (Yao *et al.,* 2017). *trnH-psbA* had the highest success rate of PCR amplification (83.87%), followed by *rbcL* (80.37%) and *matK* (59.63%), and the success rate of PCR amplification for *ITS* was the lowest (58.23%). Regarding DNA sequencing, *rbcL* and *trnH-psbA* showed the highest success rate (82.80% and 68.28%, respectively), followed by ITS (56.99%) and *matK* (50%) (Kang *et al*., 2017).The Sanger sequences results revealed that the region's sequenced were matching to the targeted chloroplastic, mitochondrial and nuclear *ITS* region and these shows that the accuracy of sanger sequencing and its potential for identification of complex flora

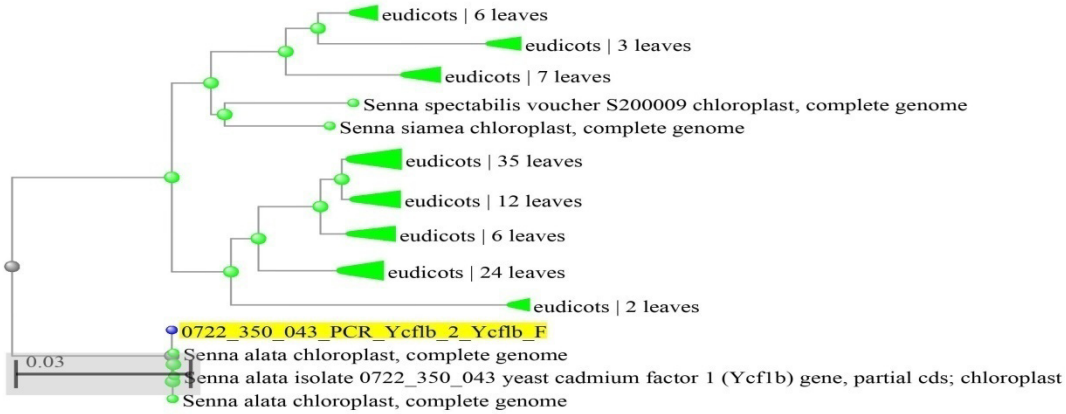

Fig. 5. The phylogenetic tree constructed for the species of *Senna alata* (*Ycf1b_2*_F)

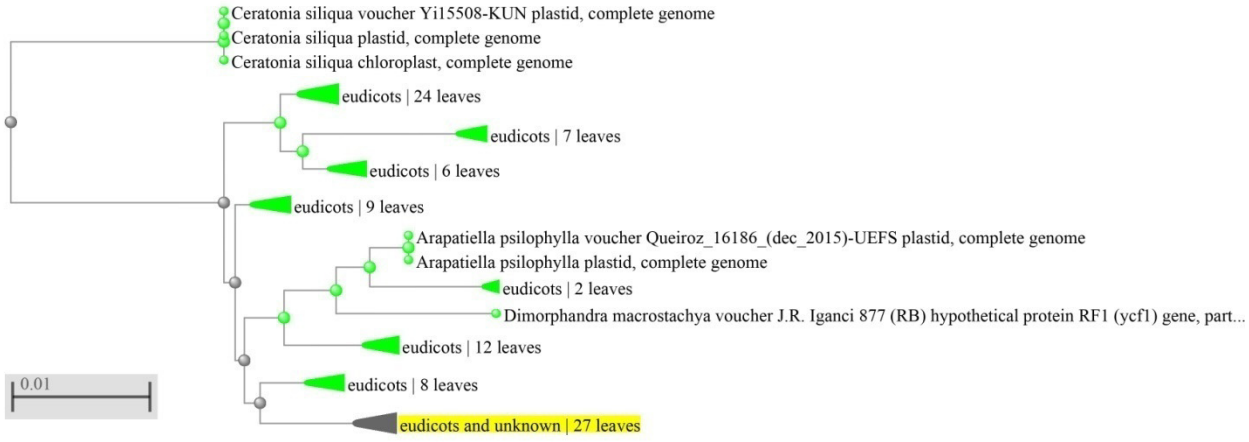

Fig. 6. The phylogenetic tree constructed for the species of Senna alata (Ycf1b_2_R)

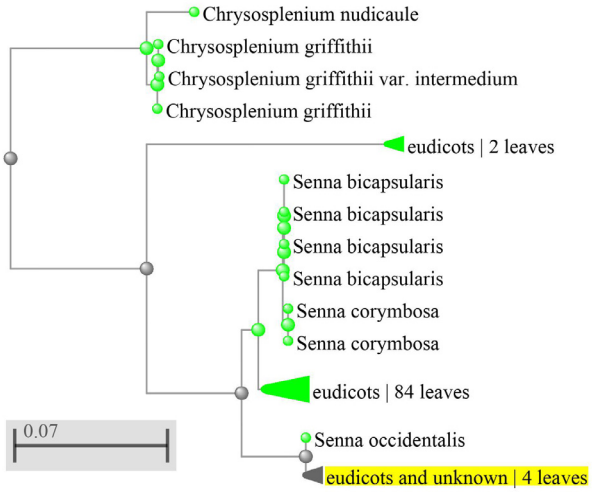

Fig. 7. The phylogenetic tree constructed for the species of *Senna occidentails (Ycf1b_3_F*

and fauna. Multiple sequence alignment from the difference species helps identify any difference in the nucleotide rearrangement such as deletion, inversion or transition. Identifying such difference in the nucleotide rearrangement could help to design single nucleotide polymorphism (*SNP*'s markers) unique to certain plant species which could play significant role in species identification (Surya and Hari, 2017). Overall the presented study proved that there is significant potential using the DNA barcode and sequencing techniques to explore more important medicinally valued intra and natural inter specific native species.

Conflict of Interest: None

References

Acharya, G.C., Mohanty, S., Dasgupta, M., Sahu, S., Singh, S., Koundinya, A.V. and Sahoo, M. R. 2022. Molecular phylogeny, DNA barcoding, and ITS2 secondary structure predictions in the medicinally important Eryngium genotypes of East Coast Region of India. *Genes*. 13(9): 1678.

Danish, M., Singh, P., Mishra, G., Srivastava, S., Jha, K.K. and Khosa, R. L. 2011. *Cassiafistula Linn. (Amulthus)-An Important Medicinal Plant: A Review of Its Traditional Uses, Phytochemistry and Pharmacological Properties*.

De Boer, H.J., Ichim, M.C. and Newmaster, S.G. 2015. DNA Barcoding and Pharmacovigilance of Herbal Medicines. *Drug Safety*. 38(7): 611–620.

De Britto, A.J., Prabha, G., Raj, T.L.S. and Kumar, P.B.J.R. 2010. genetic diversity anal genetic diversity analysis in five ysis in five ysis in five C*assia* species using rapd species using RAPD MARKERS. *Journal of Basic and Applied Biology*. 4(1&2): 162-167.

Deepak, K.V., Mohanlal, V.A., Ivin, J.J.S. and Anandan, R. 2018. DNA Barcoding of Psoraleacorylfolia, Mucunapruriens and Clitoriaternatea for Species Identification. *Int. J. Curr. Microbiol. App. Sci*. 7(12): 117-124.

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S. and Zhou, S. 2015. ycf1, the most promising plastid DNA barcode of land plants. *Scientific Reports*. 5(1): 1-5.

Habtemariam, S. 2013. Antihyperlipidemic components of *Cassia* auriculata aerial parts: identification through in vitro studies. *Phytotherapy Research*, 27(1): 152-155.

Hakkim, F.L., Girija, S., Kumar, R.S. and Jalaludeen, M.D. 2007. Effect of aqueous and ethanol extracts of *Cassiaauriculata* L. flowers on diabetes using alloxan induced diabetic rats. *Int J Diabetes Metab*. 15: 100-6.

Hubert, N. and Hanner, R. 2015. DNA Barcoding, species delineation and taxonomy: A historical perspective. *DNA Barcodes*. 3(1).

Jain, S.K. 1968. Medicinal Plants, National Book Trust, New Delhi. p. 37.

Kang, Y., Deng, Z., Zang, R. and Long, W. 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Scientific Reports*. 7(1): 1-9.

Khan, S.A., Baeshen, M.N., Ramadan, H.A. and Baeshen, N.A. 2019. ITS2: An Ideal DNA Barcode for the Arid Medicinal Plant Rhazya Stricta. *Pharmaceutical Medicine*. 33(1): 53–61.

Kress, W.J., Erickson, D.L., Swenson, N. G., Thompson, J., Uriarte, M. and Zimmerman, J.K. 2010. Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS One. 5*(11): e15409.

Mwangi, R.W., Macharia, J.M., Wagara, I.N. and Bence, R.L. 2021. The medicinal properties of *Cassia* fistula L: A review. *Biomedicine & Pharmacotherapy*, 144: 112240.

Prasath, B.H., Mohan, A.C. and Divya, S.R. 2017. Molecular characterization of medicinal plant-*Cassia* fistula L by matK gene. *International Journal for Research in Applied Science and Engineering Technology*. 5(6): 126-132.

Purushothaman, N., Newmaster, S.G., Ragupathy, S., Stalin, N., Suresh, D., Arunraj, D.R., Gnanasekaran, G., Vassou, S.L., Narasimhan, D. and Parani, M. 2014. A tiered barcode authentication tool to differentiate medicinal *Cassia* species in India. *Genetics and Molecular Research: GMR*. 13(2): 2959-2968.

Seethapathy, G.S., Ganesh, D., Santhosh Kumar, J.U., Senthilkumar, U., Newmaster, S.G., Ragupathy, S., Uma Shaanker, R. and Ravikanth, G. 2015. Assessing product adulteration in natural health products

for laxative yielding plants, *Cassia*, Senna, and Chamaecrista, in Southern India using DNA barcoding. *International Journal of Legal Medicine*. 129(4): 693-700.

Shen, Y.Y., Chen, X. and Murphy, R.W. 2013. Assessing DNA Barcoding as a Tool for Species Identification and Data Quality Control. *Plos One*. 8(2): e57125.

Singh, S., Singh, S.K. and Yadav, A. 2013. A review on *Cassia* species: Pharmacological, traditional and medicinal aspects in various countries. *AJPCT*. 3: 291–312.

Surya, S. and Hari, N. 2017. DNA barcoding of Mangroves using Ribosomal ITS Marker in Rhizophoraceae. *Int. J. Sci. Res. Sci. Technol*. 3: 606-609.

Yao, P.C., Gao, H.Y., Wei, Y.N., Zhang, J.H., Chen, X. Y. and Li, H.Q. 2017. Evaluating sampling strategy for DNA barcoding study of coastal and inland halotolerant Poaceae and Chenopodiaceae: a case study for increased sample size. *PloS One*. 12(9): e0185311.

Yu, J., Wu, X., Liu, C., Newmaster, S., Ragupathy, S. and Kress, W.J. 2021. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety*. 208: 111691.

Zahra, W., Rai, S.N., Birla, H., Singh, S. S., Rathore, A.S., Dilnashin, H., Keswani, C. and Singh, S.P. 2020. Economic Importance of Medicinal Plants in Asian Countries. In C. Keswani (Ed.). *Bioeconomy for Sustainable Development, Springer*. 359-377.