I J R B A T, Issue (XIII) Volume (II) May 2025: 30-36 e-ISSN 2347 – 517X

A Double-Blind Peer Reviewed & Refereed Journal OPEN ACCESS Original Article

INTERNATIONAL JOURNAL OF RESEARCHES IN BIOSCIENCES, AGRICULTURE AND TECHNOLOGY

COMPUTATIONAL SCREENING OF CHLOROPLAST DNA BARCODING LOCI FOR RELIABLE IDENTIFICATION OF *SPONDIAS* L. SPECIES

Stalin Nithaniyal[1]* and Jesubalan D[1]

[1]Botanical Survey of India (BSI), Western Regional Centre (WRC), Koregaon Road, Pune, Maharashtra, 411001, India.

Corresponding author: stalinbotany88@gmail.com

Communicated: 15.11.2024 Revision: 06.01.2025 & 18.01.2025 Accepted: 16.02.2025 Published: 31.05.2025

ABSTRACT:

The *Spondias* (Anacardiaceae) species has widely used in the Indian traditional system of medicine. This study aimed to evaluate the effectiveness of chloroplast DNA barcoding markers *rbcL*, *mat*K, and *trn*H-*psb*A loci to distinguish five *Spondias* species using computational approaches. We retrieved DNA barcode sequences from the NCBI and then analyzed the sequence composition using different parameters. Subsequent alignment and phylogenetic tree construction were performed to evaluate species resolution capabilities. The results indicated that the nucleotide variability improves the species resolution among the DNA barcodes. The *rbcL* and *mat*K barcodes were less suitable for distinguishing *Spondias* species. In contrast, the *trn*H-*psb*A demonstrated higher species resolution in differentiating the five species *S. dulcis*, *S. tuberosa*, *S. mombin*, *S. bahiensis*, and *S. pinnata*. The resolving power of *the trn*H-*psb*A region makes it a more suitable DNA barcode marker for assessing genetic relatedness at the species level within *Spondias*. This genetic information is critical for selecting appropriate DNA barcode loci for species authentication, particularly in the market samples of *S. pinnata*, as well as for studying the phylogenetic relationships within the Anacardiaceae family.

Keywords:- *Spondias*, Anacardiaceae, DNA barcode, *mat*K, *rbcL*, *trn*H-*psb*A, Chloroplast markers

INTRODUCTION:

The genus *Spondias* belongs to the Anacardiaceae family and is known for its medicinal, economic, and ecology value. Tropical countries including India cultivate several species for their fruits to serve as food for humans and animals. About 18 species are recognized in the *Spondias*, with nine being neotropical species. However, the taxonomy of *Spondias* is still not resolved because there are possible hybrids and species with little morphological differentiation (Silva *et al.,* 2015). Accurate taxonomic identification is crucial for conservation, breeding programs, and the sustainable use of the species. The morphological methods are often inadequate for distinguishing closely related species, particularly when dealing with morphotypes or look-alike species that resemble similar morphological features. To address this problem, DNA barcoding was widely used as a species identification tool in complex plant groups. DNA barcoding technique utilizes a short, standardized DNA segment from a uniform locality of the genome to identify species. Therefore, this technique offers a rapid and accurate method for species identification and to distinguish closely related species (Hebert *et al.,* 2003). Plant DNA barcoding research commonly adopts the chloroplast DNA regions *rbc*L, *mat*K, and the *trn*H-*psb*A intergenic spacer as barcoding markers due to their relatively high sequence conservation and the availability of universal primers (CBOL, 2009). Earlier studies reported the effectiveness of chloroplast markers in providing high-quality sequences for species-level identification across a diverse range of plant taxa

http://doi.org/10.29369/ijrbat.2025.010.2.0003

(Hollingsworth *et al.,* 2011). However, species resolution could vary in the different plant groups and the effectiveness of group-specific barcoding markers within the *Spondias* remains unestablished. Several research has shown that the *rbc*L and *mat*K regions are useful for barcoding in some plant families, but their discriminatory power may be limited in others (Saarela *et al.,* 2013). Despite its non-coding nature, DNA barcoding studies have found the *trn*H-*psb*A region to have higher discrimination power in certain taxa, making it a valuable DNA barcode complement to the *rbc*L and matK regions (Pang *et al.,* 2011). The current study aims to conduct a computational evaluation of the chloroplast DNA barcoding markers *rbcL*, *mat*K, and *trn*H-*psb*A in the *Spondias* genus. This study seeks to assess the efficacy of *rbcL*, *mat*K, and *trn*H-*psb*A markers in species identification and phylogenetic resolution within the *Spondias*. The findings will provide novel insights into the genetic relationships among *Spondias* species and provide reliable DNA barcoding protocols to distinguish complex genera in species identification, authentication of herbal drugs, and conservation of species.

METHODS :

Dataset of *Spondias* species

The DNA sequences of *rbcL*, *mat*K, and *psb*A-*trn*H of five *Spondias* species belonging to the Anacardiaceae family were retrieved from the chloroplast genome sequence from the nucleotide database of NCBI (http://www.ncbi.nlm.nih.gov). All of the sequences were downloaded from the respective chloroplast genome sequences and the species details with GenBank accession number are *Spondias dulcis* (MZ929415), *Spondias tuberosa* (KU756562), *Spondias mombin* (KY828469), *Spondias bahiensis* (KU756561), and *Spondias pinnata* (OP650214).

Validation of DNA barcode sequences

The GenBank database of NCBI was searched for the presence of *Spondias* DNA barcode sequences. The DNA sequences were validated using the criteria as follows i) the sequence was tagged with publication in refereed journals, ii) placed in a clade that is relative to its taxonomic affiliation, and iii) showed correct sequence alignment using the MUSCLE algorithm with 70% threshold (Sivaraj *et al.* 2018; Yadav *et al.* 2024).

Computational analysis of DNA barcoding markers

The FASTA formatted sequence dataset was prepared for each species for the computational analysis. The nucleotide composition, AT and GC content, and the average size of the aligned length were calculated. The sequences were assembled for Multiple Sequence Alignment (MSA) using ClustalW by using MEGA 11 software (Tamura *et al.,* 2021). Evolutionary divergence and nucleotide substitution pattern analysis were performed. The maximum likelihood tree method was used for phylogenetic analysis with a 1000 bootstrap value. Species resolution was estimated for a given barcode locus by considering the species of conspecific individuals who were grouped in a monophyletic clade with high (>70%) bootstrap support.

RESULTS AND DISCUSSION :

Analysis of genetic distance in *Spondias*

The size variation in the sequence was not observed between species for the *rbc*L (1428 bp) and *mat*K (1518 bp) DNA barcode markers. Whereas the *trn*H-*psb*A non-coding region displayed size variations ranging from 530 bp to 598 bp. The genetic distance in the *rbc*L, *mat*K, and *trn*H-*psb*A regions was found to be 0.00-0.01, 0.00-0.02, and 0.01-0.06, respectively

(Tables 1, 2, and 3). The *rbc*L region of *S. mombin* showed slightly more genetic distance than the other *Spondias* species, with a p-distance of 0.01. Similarly, the *mat*K region of *S. mombin* showed a significant genetic distance, with a p-distance ranging from 0.01 to 0.02. In the *trn*H-*psb*A region, all five *Spondias* species showed the highest genetic distance, ranging from 0.01 to 0.06. This finding aligns with a previous study, which demonstrated that the *trn*H-*psb*A region had significantly higher identification efficiency than *mat*K and *rbc*L across 18 families and 21 genera (Pang *et al.,* 2011). Moreover, recently reported low discrimination efficiency of *rbc*L and *mat*K in a study involving 16 Anacardiaceae species (Ryadi *et al.,* 2023).

Table 1: Estimation of evolutionary distance in *rbcL* sequences of *Spondias* taxa

	Species name	1	2	3	4	5
1	*Spondias bahiensis*	0.00				
2	*Spondias dulcis*	0.01	0.00			
3	*Spondias mombin*	0.01	0.01	0.00		
4	*Spondias tuberosa*	0.00	0.00	0.01	0.00	
5	*Spondias pinnata*	0.01	0.00	0.01	0.01	0.00

Table 2: Estimation of evolutionary distance in *matK* sequences of *Spondias* taxa

	Species name	1	2	3	4	5
1	*Spondias bahiensis*	0.00				
2	*Spondias dulcis*	0.01	0.00			
3	*Spondias mombin*	0.01	0.01	0.00		
4	*Spondias tuberosa*	0.00	0.01	0.01	0.00	
5	*Spondias pinnata*	0.01	0.00	0.02	0.01	0.00

Table 3: Estimation of evolutionary distance in *trnH-psbA* sequences of *Spondias* taxa

	Species name	1	2	3	4	5
1	*Spondias bahiensis*	0.00				
2	*Spondias dulcis*	0.03	0.00			
3	*Spondias mombin*	0.04	0.01	0.00		
4	*Spondias tuberosa*	0.05	0.04	0.04	0.00	
5	*Spondias pinnata*	0.01	0.04	0.05	0.06	0.00

Nucleotide Substitution Analysis

Nucleotide substitution analysis, which includes both transition and transversion events at codon positions, provides insights into evolutionary trends. In this study, substitution patterns were evaluated across all codon positions (1st, 2nd, and 3rd nucleotides) (Table 4). Generally, transitional substitutions were more frequent than transversion substitutions in both coding regions (*rbc*L and *mat*K) and the non-coding region (*trn*H-*psb*A). The *rbc*L region exhibited a higher substitution rate for T to C and A to G compared to the *mat*K and *trn*H-*psb*A regions. Conversely, the substitution frequency from G to A and C to T was higher in the *trn*H-*psb*A region than in *rbc*L and *mat*K.

Phylogenetic Analysis and evolutionary relationship within *Spondias*

The resolution of a DNA barcode is comprehended with the ability to differentiate and identify species based on interspecific differences in DNA sequences. A species is considered resolved if its individuals form a specific monophyletic clade (Nithaniyal *et al.,* 2014). Using this criterion, three out of the five species, *S. dulcis*, *S. mombin*, and *S. pinnata*, were resolved in the *rbc*L and *mat*K regions. For clarity, the Phylogenetic tree with bootstrap values >80% is shown in the figures (Figures 1 and 2). The phylogenetic trees of *rbc*L and *mat*K DNA barcode did not resolve the species of *Spondias* species. In contrast, the *trn*H-*psb*A region successfully resolved all five species, demonstrating high discrimination efficiency. To further enhance the resolving ability of these three regions, additional studies with larger sequence datasets are needed. A study on DNA barcoding of five *Spondias* species revealed that *matK* and *rbc*L genes were not effective for species discrimination due to their low discriminatory power, while *trn*H-*psb*A exhibited a high level of discrimination. However, the two species *S. venulosa* and *S. tuberosa* were differentiated with albeit low divergence (Silva *et al.,* 2015). In this study, neither *rbc*L nor *mat*K could discriminate all species but, the *trn*H-*psb*A sequence discriminated all species in the phylogenetic analysis, and the results were presented in Figure 2. The phylogenetic tree of *trn*H-*psb*A sequences resolved five species such as *S. dulcis*, *S. tuberosa*, *S. mombin*, *S. bahiensis*, and *S. pinnata*. Several studies, including those on *Paphiopedilum* (Vu *et al.,* 2019), *Araucaria* and Inga (Hollingsworth *et al.,* 2009), *Terminalia* species (Nithaniyal and Parani, 2016), and temperate flora (Burgess *et al.,* 2011), have reported the high capacity of the *trn*H-*psb*A DNA barcode region in plant identification. However, its discrimination efficiency may be lower in some cases, such as with tropical tree species (Gonzalez *et al.,* 2009). The effectiveness of using *rbc*L, *mat*K, or *trn*H-*psb*A alone for plant species discrimination is limited and in such cases, additional markers like ITS2 could be employed for unambiguous identification (Yao et al., 2010).

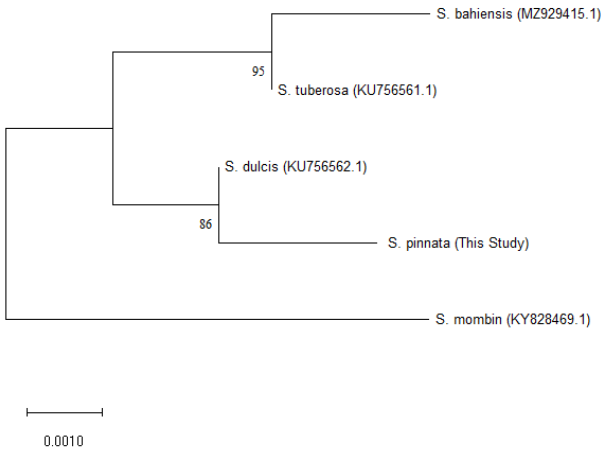

Figure 1. Maximum Likelihood tree of *rbcL* DNA barcode sequences of *Spondias*. Numbers above branches denotes the bootstrap values.

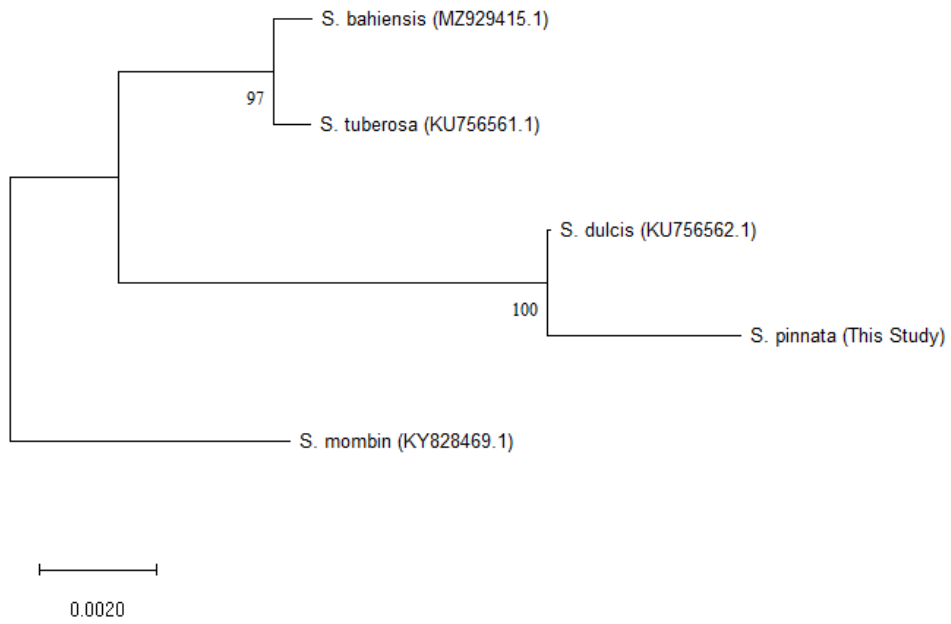

Figure 2. Maximum Likelihood tree of *matK* DNA barcode sequences of *Spondias*. Numbers above branches denotes the bootstrap values.

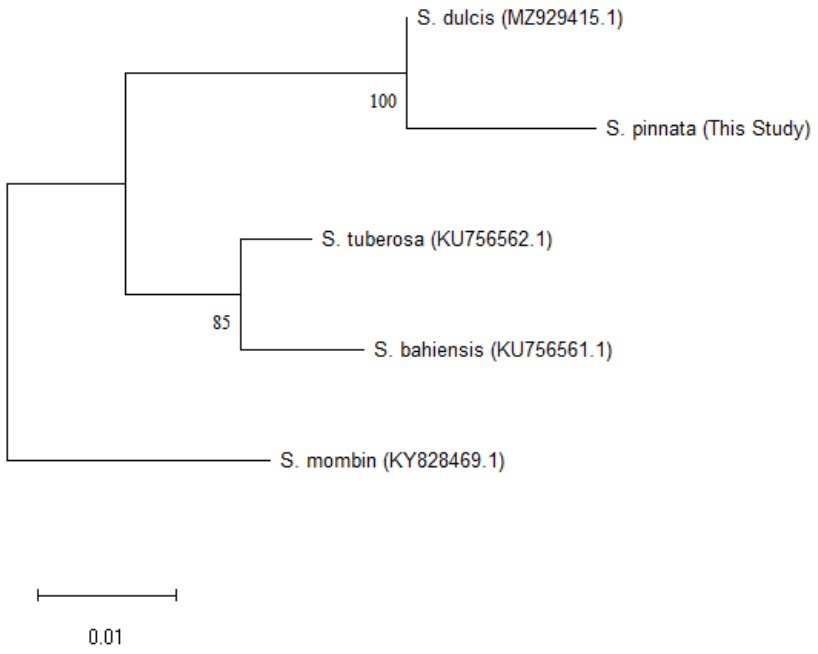

Figure 3. Maximum Likelihood tree of *trnH-psbA* DNA barcode sequences of *Spondias*. Numbers above branches denotes the bootstrap values.

CONCLUSION

The results of this study suggest that the species-resolving ability of the *rbc*L, *mat*K, and *trn*H-*psb*A regions varied among different *Spondias* species. The non-coding *trn*H-*psb*A barcode significantly enhances species discrimination. However, the limited availability of sequences for some species may affect the precision of the study. Therefore, further comparative chloroplast genome studies should include with other coding regions to identify the potential barcode marker among the chloroplast genes and to improve species identification, discrimination. This approach could enhance the effectiveness of DNA barcoding for *Spondias* species within the Anacardiaceae family.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest related to this article.

Acknowledgements

The authors are thankful the Director, Botanical Survey of India and the Head of office, Western Regional centre, Botanical Survey of India, Pune, India

REFERENCES:

Burgess, K.S., Fazekas, A.J., Kesanakurti, P.R., Graham, S.W., Husband, B.C., Newmaster, S.G., et al. Discriminating Plant Species in a Local Temperate Flora Using the rbcL+matK DNA Barcode. Methods in Ecology and Evolution, 2011, 2, 333-340.

Gonzalez MA, Baraloto C, Engel J, Mori SA, Pe´tronelli P, et al. Identification of Amazonian trees with DNA barcodes. PLoS ONE, 2009, 4(10): e7483.

CBOL Plant Working Group 1, Hollingsworth, P. M., L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham and D. P. Little. A DNA barcode for land plants. Proceedings of the National Academy of Sciences, 2009, 106(31), 12794-12797.

Hebert, P. D., S. Ratnasingham and J. R. De Waard. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences, 2003, 270(suppl_1), S96-S99.

Hollingsworth, M. L., Andra Clark, A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., Cowan, R., Chase, M. W., Gaudeul, M., & Hollingsworth, P. M. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular ecology resources, 2009, 9(2), 439–457.

Hollingsworth, P., Graham, S., Little, D.,. Choosing and using a plant DNA barcode. PLoS one, 2011 6, e19254.

Nithaniyal S, Parani M. Evaluation of chloroplast and nuclear DNA barcodes for species identification in *Terminalia* L. Biochemical Systematics and Ecology. 2016 1;68:223-9.

Nithaniyal, S., S. G. Newmaster, S. Ragupathy, D. Krishnamoorthy, S. L. Vassou and M. Parani. DNA barcode authentication of wood samples of threatened and commercial timber trees within the tropical dry evergreen forest of India. PLoS One, 2014, 9(9), e107669.

Pang, X., Song, J., Zhu, Y., Xu, H., Huang, L., and Chen, S. Applying plant DNA barcodes for Rosaceae species identification. Cladistics, 2011, 27: 165–170. doi: 10.1111/j.1096-0031.2010.00328.x

RyadiA., SiregarI., MouraC. de M., GailingO., & AmanditaF. Y. An Early Reference to DNA Barcode for the Anacardiaceae

Family. HAYATI Journal of Biosciences, 2023, 30(3), 543-550.

Saarela JM, Sokoloff PC, Gillespie LJ, Consaul LL, Bull RD. DNA Barcoding the Canadian Arctic Flora: Core Plastid Barcodes (*rbcL* + *matK*) for 490 Vascular Plant Species. PLoS ONE, 2013, 8(10): e77982.

Silva JN, da Costa AB, Silva JV, Almeida C. DNA barcoding and phylogeny in neotropical species of the genus *Spondias*. Biochemical Systematics and Ecology 2015, 61:240-243. DOI: 10.1016/j.bse.2015.06.005.

Sivaraj, I., S. Nithaniyal, V. Bhooma, U. Senthilkumar, and M. Parani. Species delimitation of *Melia dubia* Cav. from *Melia azedarach* L. complex based on DNA barcoding. Botany, 2018, 96(5), 329-336.

Tamura, K., G. Stecher and S. Kumar. MEGA11: molecular evolutionary genetics analysis version 11. Molecular biology and evolution, 2021, 38(7), 3022-3027.

Vu, H. T., Vu, Q. L., Nguyen, T. D., Tran, N., Nguyen, T. C., Luu, P. N., Tran, D. D., Nguyen, T. K., & Le, L. Genetic Diversity and Identification of Vietnamese *Paphiopedilum* Species Using DNA Sequences. Biology, 2019, 9(1), 9.

Yadav, P., S. Nithaniyal, S. D. Mishra and M. P. Goutam. Forensic application of DNA barcoding to differentiate toxic plant from the common medicinal plants for consumer safety. African Journal of Biological Sciences, 2024, 6:1–10.

Yao, H., J. Song, C. Liu, K. Luo, J. Han, Y., Li, X. Pang, H. Xu, Y. Zhu P. Xiao and S. Chen. Use of ITS2 region as the universal DNA barcode for plants and animals. PloS one, 2010, 5(10), e13102.