Journal of AgriSearch, 12 (2): 81-86
An Open Access International Peer Reviewed Quarterly

ISSN : 2348-8808 (Print), 2348-8867 (Online)
https://doi.org/10.21921/jas.v12i02.15243

In Silico Analysis of a Chloroplast Loci for DNA Barcoding of Aloe spp.

Sweta Gupta[1] and Punam Sinha[2*]

ABSTRACT

DNA barcoding has emerged as a reliable tool for species identification, especially in taxonomically complex plant groups such as the genus Aloe. This study employed an in-silico approach to evaluate the efficacy of the rbcL gene locus as a DNA barcode for 20 Aloe species. A total of 70 high-quality sequences linked to voucher specimens were retrieved from GenBank, manually curated to a uniform length of 480 base pairs, and analyzed using multiple sequence alignment and phylogenetic methods. The evolutionary divergence within and among species was estimated using p-distance metrics, revealing high sequence conservation across most Aloe species. Phylogenetic analysis using the Neighbor-Joining method demonstrated that 12 species (60%) formed well-supported monophyletic clades, confirming the rbcL locus's moderate resolution capacity. Species such as A. vera, A. lineata, and A. maculata clustered tightly with high bootstrap values, while few closely related species remained unresolved. The findings suggest that the sample size and incorporating multi-locus analyses will further improve species identification and evolutionary understanding within the genus.

Keywords: DNA barcoding, Aloe species, RBCL gene, Phylogenetic analysis, Species resolution, In silico study

ARTICLE INFO

Received on	:	29/05/2025
Accepted on	:	18/06/2025
Published online	:	30/06/2025

CC BY SA

INTRODUCTION

DNA barcoding has exhibited tremendous implication in species identification and biodiversity assessment across various biological domains. It involves the use of short, standardized gene regions as molecular tags to identify and discriminate between species. Among the candidate loci for plant DNA barcoding, the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene (rbcL) has been extensively employed due to its universality, moderate evolutionary rate, and availability of universal primers (Kress and Erickson, 2007).

The rbcL gene, which encodes the large subunit of the enzyme RuBisCO, is located in the chloroplast genome and is considered one of the most conserved protein-coding genes in plants. This attribute makes rbcL a foundational barcode for initial taxonomic assignments, particularly when used in combination with faster-evolving loci like matK or ITS.

The genus Aloe (family: Asphodelaceae) comprises over 500 species distributed primarily in Africa, the Arabian Peninsula, and Madagascar. These species are economically and medicinally significant, with various applications in pharmaceuticals, cosmetics, and traditional medicine (Grace et al., 2015). However, morphological plasticity, hybridization, and limited reproductive barriers among Aloe species pose significant challenges in their accurate identification using classical taxonomy. Accurate identification is critical not only for biodiversity inventories but also for conservation management and regulation of commercially valuable species like Aloe vera. In this regard, DNA barcoding offers a robust mechanism for species delimitation, particularly when traditional approaches fall short.

Several large-scale initiatives, including the Barcode of Life project and the CBOL Plant Working Group, have recommended rbcL as one of the core barcode regions for plants. Despite its limitations in distinguishing closely related species, its universality and ease of amplification remain unmatched (CBOL Plant Working Group, 2009). In genera with limited molecular data, such as Aloe, rbcL provides a starting point for constructing DNA barcode libraries that can be later refined with additional markers.

In recent years, in silico approaches have gained prominence in the evaluation and optimization of barcode loci. These computational methods facilitate large-scale data analysis, sequence alignment, and phylogenetic inference without extensive laboratory experimentation (Little, 2011).

The in silico approach adopted in this study provides multiple

[1]Assistant Professor, Department of Botany, Ganaga Devi Mahila Mahavidyalaya, Patliputra University, Patna
[2]Head, Department of Botany, Ganaga Devi Mahila Mahavidyalaya, Patliputra University, Patna
[*]Corresponding author email: punam.gdmc@gmail.com

advantages. By utilizing bioinformatics platforms such as BLAST, ClustalW and MEGA barcode loci across numerous accessions can be evaluated without the need for wet-lab experimentation. Sequence retrieval from GenBank and BOLD ensures access to a wide range of geographically and taxonomically diverse samples, thereby increasing the robustness of analytical outputs.

Moreover, in silico analyses allow for simulation-based evaluation of discriminatory power using tree-based and distance-based methods. Phylogenetic tree construction provides a visual representation of species clusters, enabling the identification of monophyletic clades and assessment of barcode resolution (Hebert et al., 2003).

The rbcL gene has been widely used in plant barcoding studies, but its efficacy varies across different taxonomic groups. In some plant lineages, rbcL alone does not provide sufficient resolution at the species level, necessitating its combination with other loci such as matK or ITS (CBOL Plant Working Group, 2009). Therefore, it is essential to evaluate the performance of rbcL specifically within the Aloe genus using comprehensive in silico analyses.

The present study aims to assess the effectiveness of the rbcL locus for DNA barcoding of Aloe spp. using an in silico approach. The objectives include sequence analysis of rbcL sequences of Aloe species from NCBI in order to evaluate species resolution, and determining the potential of rbcL as a reliable barcode marker for Aloe species identification.

MATERIALS AND METHODS

DNA sequences of the rbcL gene corresponding to voucher specimens of Aloe species were included in this study. A total of 20 Aloe species, each represented by at least two available accessions, were retrieved from GenBank and downloaded in FASTA format. Unverified sequences, as well as those containing more than 3% ambiguous nucleotides, were excluded from the analysis following the standard protocol (Suesatpanit et al., 2017). The selected sequences were manually curated to ensure a uniform length of 480 base pairs (positions 98–577) for consistent comparative analysis. Multiple sequence alignment was performed using the ClustalW algorithm implemented in MEGA 12 software (Kumar et al., 2024). Subsequent analyses, including nucleotide substitution modeling, genetic distance estimation, and phylogenetic inference, were also conducted using the same software. Phylogenetic analysis was performed using the Neighbor joining method for nucleotide substitution. A phylogenetic tree was constructed, and species resolution efficacy was assessed based on the clustering pattern. Species were considered resolved if their accessions formed a distinct monophyletic clade with strong bootstrap support. Conversely, species with accessions dispersed across paraphyletic branches were considered unresolved, indicating failure of accurate identification (Sikdar et al., 2018)

Table 1: Aloe species under study with their sequence detail

Aloe Species	Gene bank Accession number of the sequences used in the study	No of rbcL gene sequence selected	No of rbcL gene sequence downloaded
Aloe aageodonta	KU748219.1 KU748277.1	02	02
Aloe lineata	JQ024519.1,JQ024520.1 , JQ024521.1, JQ024522.1	04	06
Aloe knifophiades	JX572285.1, KC960550.1		
Aloe powysiorum	KU748279.1 KU748334.1, KU748336.1, KU748343.1 , KU748349.1,KU748197.1	06	06
Aloe diolii	KU748389.1,KU748390.1 KU748391.1	03	04
Aloe elgonica	KU748228.1,KU748276.1	02	02
Aloe buhrii	JQ024494.1,KU748384.1, KU748385.1, KU748386.1	04	05
Aloe purpurea	KX270420.1,KX270422.1	02	11
Aloe turkanensis	KU748067.1, MT231436.1, KU748339.1	03	03
Aloe vera	NC035506.1, ON641334.1,ON641362.1, ON651207.1,OQ392338.1, KP072713.1 KP072714.1, KP072715.1, KP072716.1 KP072717.1, KP072718.1, JQ273907.1 MT385749.1,MW091969.1	14	34
Aloe striata	JQ024533.1,JQ024534.1	02	05

Aloe Species	Gene bank Accession number of the sequences used in the study	No of rbcL gene sequence selected	No of rbcL gene sequence downloaded
Aloe secundiflora	MT231385.1, MT231386.1, MT231433.1, MT231434.1	04	04
Aloe rivae	KU748076.1,KU748199.1	02	02
Aloe maculate	KP072708.1,KP072709.1,KP072710.1, KX377523.1, NC035505.1, JQ412311.1	06	10
Aloe confusa	KU748275.1,KU748387.1 KU748287.1	03	04
Aloe comosa	JQ024498.1, JQ024499.1	02	02
Aloe nyrensis	KU748064.1KU748200.1,MT231435.1	03	07
Aloe brevifolia	JX572275.1,JQ024493.1	02	03
Aloe chrysostachuys	KU748225.1,KU748226.1	02	02
Aloe melanacantha	JQ024523.1, JX572287.1	02	03

RESULTS AND DISCUSSION

The search using key word *Aloe+rbcL* yielded multiple sequences across different species mentioned in of table 1. A total of 70 sequences belonging to 20 *Aloe* species fulfilling the inclusion criteria of the study, were found suitable for further analysis. Only sequences linked to voucher specimens were considered to enhance the reliability of species-level identification, which limited the dataset to 20 species and 70 curated sequences. The selection of 480 sequence length of *rbcL* genes was aimed to get result with accuracy. Notably, *Aloe vera* had the highest representation with 34 sequences available, of which 14 species fulfilling criteria above were selected for analysis.

Estimation of Sequence Divergence

The variation among 20 *Aloe* species under study as well as within individuals of a species was determined and presented in table-2 &3 as evolutionary divergence (p distance value). Out of all species examined, 17 species showed no variations among individuals. The lack of variations within species may be attributed to small number of individual samples (2-3 sequences) selected for analysis but, interestingly *Aloe vera* showed no variations despite having 14 individual sequences under study. The variations observed among three individuals ranged between 0.003 to 0.083 in the species *Aloe turkanensis, Aloe secundiflora,* and *Aloe burhii*, which is comparatively lesser in comparison to similar type of study (Ho and Nguyen, 2020). The aforesaid finding suggests that the *rbcL* gene is highly conserved in *Aloe* species and can be of taxonomical importance. The *rbcL* gene based variability among *Aloe* species varied from 0.000 to 0.060 (Tables 2) with no interspecies variation of in few species such as *Aloe diolii* to *Aloe powysiorum/Aloe confuse/Aloe chrysostachuys/Aloe aageodonta/Aloe rivae* and *Aloe powysiorum* to *Aloe confuse/Aloe aageodonta/Aloe chrysostachuys/Aloe rivae* which reveals presence of uniformity in evolution among these species.

Estimation of Species Resolution

The resolution capacity of a DNA barcode reflects its effectiveness in differentiating species based on interspecific sequence variation. A species is considered resolved when all its individuals cluster into a well-supported monophyletic group (Sikdar *et al*., 2018). The present study successfully differentiates 12 out of 20 Aloe species, includes *Aloe brevifolia, Aloe chrysostachuys, Aloe vera, Aloe confuse, Aloe comosa, Aloe maculate, Aloe lineata, Aloe diolii, Aloe elgonica, Aloe purpurea, Aloe striata* and *Aloe melanacantha* based on the phylogenetic tree constructed (Fig. 1) which indicates that *rbcL* gene has 60% species resolution ability for *Aloe* species. The bootstrap values below 50% have been excluded from the figure for clarity. *Aloe vera* clustered together with high bootstrap support, reinforcing their close evolutionary relationship. This aligns with earlier findings by Treutlein *et al*. (2003), who demonstrated genetic similarity among these morphologically alike species. *Aloe knipholioides* and *A. striata* also clustered in a moderately supported group, suggesting evolutionary proximity, although the bootstrap value (50%) suggests moderate confidence. The tree displays four or more sequences for certain species, such as *A. lineata* and *A. maculate* with intra-species sequences clustered tightly, indicating the *rbcL* locus is consistent within species and has minimal intraspecific variation. Conversely, the inter-species divergence is clearly visible across distinct clades, supporting the role of *rbcL* in species-level discrimination. This observation is consistent with results from Bafeel *et al.* (2012), who reported that *rbcL* could differentiate most monocotyledonous taxa at the genus level, and in some cases, at the species level. Bootstrap values vary across the tree, with several clades showing high statistical support (≥80%), indicating reliable clustering. However, some internal branches and basal nodes demonstrate low support (<60%),

suggesting ambiguity in deeper evolutionary relationships. Similar observations were reported by Roy *et al*. (2010) in their molecular phylogeny studies on *Aloe*, where deeper nodes often required multi-locus datasets for robust resolution. The similar study carried out for family cucurbitaceae resulted into only 35 % resolving efficiency based on *matk* gene which was increased to 60% when analysis was carried out by combining *rbcL* gene with *matk* genes. The variable results have been obtained in a study of other *rbcL* loci. In a study involving 17 cultivars and species of the genus *Prunus*, demonstrated that the *rbcL* gene is an effective marker for analyzing relationships both within and among *Prunus* species (Sarhan *et al*. 2016). On the contrary, phylogenetic study of 16 species of *Setaria* genus shows that the *rbcL* gene is highly conserved at the interspecies level and is not able to differentiate species of *Setaria* genus. The present study includes a relatively high sample size of 14 sequences for *Aloe vera* compared to six each for *A. maculate/A powysiorum*, four for *Aloe lineata/ Aloe buhrii/ Aloe secundiflora* and three each for *Aloe confusa, Aloe turkanensis, and A. nyeriensis*, while the remaining species are represented by two samples each. Despite this variation in sample size, the rbcL gene-based analysis demonstrated strong potential in resolving species-level relationships. However, the availability of *rbcL* sequences in the NCBI database remains limited for certain species and increasing the number of samples is essential for obtaining more reliable insights into the species resolution capacity of this marker, as also observed by Sikdar *et al*. (2018). Although, *rbcL* performs adequately for broad phylogenetic grouping, supplementation with other loci such as *matK* or *ITS2* may be necessary to increase discrimination power for taxonomically complex species.

Table 2: Average evolutionary Divergence of rbcL regions

Species	p	SE
Aloe vera	0	0
Aloe turkanensis	0.083	0.010
Aloe nyrensis	0	0
Aloe maculata	0	0
Aloe secundiflora	0.047	0.007
Aloe purpurea	0	0
Aloe diolii	0	0
Aloe burhii	0.003	0.001
Aloe powysiorum	0	0
Aloe confusa	0	0
Aloe elongica	0	0
Aloe chrysostachuys	0	0
Aloe aageodonta	0	0
Aloe rivae	0	0
Aloe knifophiades	0	0
Aloe melanacantha	0	0
Aloe brevifolia	0	0
Aloe striata	0	0
Aloe lineata	0	0
Aloe comosa	0	0

Table 3: Evolutionary divergences in rbcL sequence pairs between 20 species of Aloe genus

Aloe vera																			
Aloe turkanensis	0.054																		
Aloe nyrensis	0.002	0.052																	
Aloe maculata	0.0104	0.054	0.008																
Aloe secundiflora	0.0255	0.060	0.024	0.027															
Aloe purpurea	0.004	0.054	0.002	0.010	0.025														
Aloe diolii	0.002	0.052	0.000	0.008	0.024	0.002													
Aloe burhii	0.003	0.053	0.001	0.007	0.024	0.003	0.001												
Aloe powysiorum	0.002	0.052	0.000	0.008	0.024	0.002	0.000	0.001											
Aloe confusa	0.002	0.052	0.000	0.008	0.024	0.002	0.000	0.001	0.000										
Aloe elongica	0.004	0.052	0.002	0.006	0.025	0.004	0.002	0.003	0.002	0.002									
Aloe chrysostachuys	0.002	0.052	0.000	0.008	0.024	0.002	0.000	0.001	0.000	0.000	0.002								
Aloe aageodonta	0.002	0.052	0.000	0.008	0.024	0.002	0.000	0.001	0.000	0.000	0.002	0.000							
Aloe rivae	0.002	0.052	0.000	0.0083	0.0240	0.0021	0.0000	0.0016	0.0000	0.0000	0.0021	0.0000	0.0000						

Aloe knifophiades	0.0063	0.0542	0.0042	0.004	0.025	0.006	0.004	0.003	0.004	0.004	0.006	0.004	0.004	0.004					
Aloe melanacantha	0.006	0.054	0.004	0.008	0.026	0.006	0.004	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.004				
Aloe brevifolia	0.006	0.054	0.004	0.008	0.026	0.006	0.004	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.004	0.004			
Aloe striata	0.006	0.054	0.004	0.004	0.025	0.006	0.004	0.003	0.004	0.004	0.006	0.004	0.004	0.004	0.000	0.004	0.004		
Aloe lineata	0.006	0.054	0.004	0.008	0.026	0.006	0.004	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.004	0.004	0.004	0.004	
Aloe comosa	0.004	0.052	0.002	0.006	0.024	0.004	0.002	0.002	0.002	0.002	0.004	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

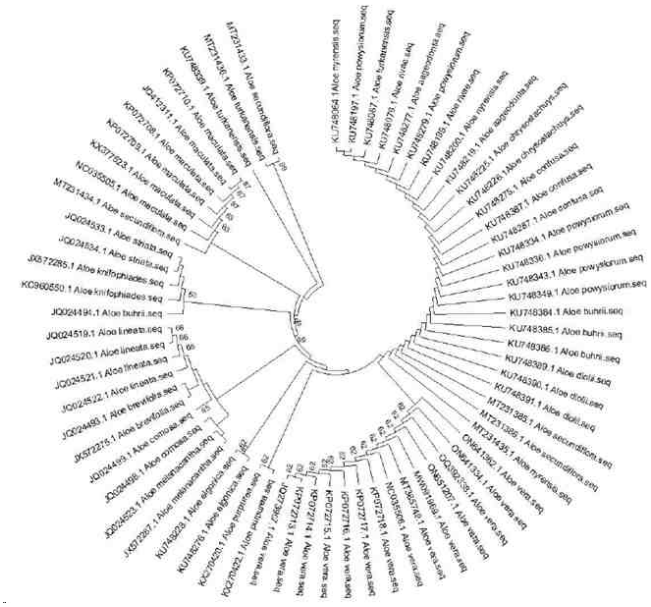

Fig. 1: Neighbor joining with 1000 bootstrap replicated based on rbcL sequences

CONCLUSION

The present analysis demonstrates that the rbcL gene is a valuable marker for the preliminary phylogenetic assessment and DNA barcoding of Aloe spp., particularly for distinguishing well-separated taxa and verifying taxonomic placements. Its universality, and alignment ease make it suitable for standard barcoding purposes. However, the limited variability of rbcL reduces its resolution power for very closely related species. To enhance the discriminatory capability and robustness of phylogenetic inference, future studies should adopt a multi-locus barcoding approach incorporating matK, trnH-psbA, and ITS2 regions. The present study has small number of high quality sequences available for analysis, therefore, future efforts should focus on expanding the reference library of high-quality, voucher-linked sequences and adopting multilocus barcoding strategies for more robust taxonomic assessments in Aloe and other complex plant genera.

ACKNOWLEDGEMENT

The authors acknowledges Dr. Bhavya Jha, Assistant Professor, Department of Zoology, GDMM, Patliputra University, Patna for scientifically fruitful discussions.

REFERENCES

Bafeel S O, Arif I A, Bakir M A, Al Homaidan A A, Al Farhan A H, Khan H A. 2012. DNA barcoding of arid wild plants using rbcL gene sequences. Genet Mol Res. 11(3):1934-41.PMID: 22869548. https://doi.org/10.4238/2012.july.19.12

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106(31): 12794–12797. https://doi.org/10.1073/pnas.0905845106

Grace O M, Buerki S, Symonds M R, Forest F, van Wyk A E, Smith G F and Klopper R R. 2015. Evolutionary history and leaf succulence as explanations for medicinal use in aloes and the global popularity of Aloe vera. BMC Evolutionary Biology 15(1): 29. https://doi.org/10.1186/s12862-015-0291-7

Hebert P D N, Cywinska A, Ball S L and deWaard J R. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270(1512): 313–321. https://doi.org/10.1098/rspb.2002.2218

Ho V T and Nguyen P M. 2020. An in silico approach for evaluation of rbcl and matk loci for DNA barcoding of cucurbitaceae family. Biodiversitas 21(08): 3879-3885. http://dx.doi.org/10.13057/biodiv/d210858

Kress W J and Erickson D L. 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One 2(6): e508. https://doi.org/10.1371/journal.pone.0000508

Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, and Tamura K. 2024. MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. Molecular Biology and Evolution 41: 1–9 https://doi.org/10.1093/molbev/msae263

Little D P. 2011. DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. PLoS One 6(8): e20552. https://doi.org/10.1371/journal.pone.0020552

Roy S, Tyagi A, Shukla V, Kumar A, Singh U M, Chaudhary LB, Datt B, Bag S K and Singh P K. 2010. Universal plant DNA barcode loci may not work in complex

groups: a case study with Indian berberis species. PLoS ONE 5(10): e13674. https://doi.org/10.1371/journal.pone.0013674

Sarhan S, Hamed F and Al-Youssef W. 2016. The rbcL Gene Sequence Variations among and within Prunus Species. Journal of Agricultural Science and Technology 18(4):1105-15.

Sikdar S, Tiwari S, Sapre S and Thakur V V. 2018. Simple approach for species discrimination of fabaceae family on the basis of length variation in PCR amplified products using barcode primers. International Journal of Current Microbiology and Applied Science 7: 921-8. https://doi.org/10.20546/ijcmas.2018.712.115

Suesatpanit T, Osathanunkul K, Madesis P, Osathanunkul M. 2017. Should DNA sequence be incorporated with other taxonomical data for routine identifying of plant species? BMC ComplAltern Med 17: 437. https://doi.org/10.1186/s12906-017-1937-3

Treutlein J, Smith G F, van Wyk B E and Wink M. 2003. Phylogenetic relationships in Aloe (Asphodelaceae) inferred from chloroplast DNA sequences (rbcL, matK) and AFLP data. Plant Systematics and Evolution 239(1-2): 141-153.

Citation:
Gupta S and Sinha P.2025. In Silico analysis of a chloroplast loci for DNA Barcoding of Aloe spp. Journal of AgriSearch 12(2):81-86.