Indian Journal of Plant Sciences ISSN: 2319–3824
Online, International Journal, Available at http://www.cibtech.org/jps.htm
2025 Vol.14, pp.54-61/Sweta et al.
Research Article (Open Access)

EVALUATION OF MATK LOCI FOR DNA BARCODING OF ALOE SPP: AN *IN SILICO* STUDY

Sweta Gupta[1], Soni Kumari[2] and *Punam Sinha[1]
[1]*Department of Botany, Ganga Devi MahilaMahavidyalaya, Patliputra University, Patna-800020, Bihar*
[2]*Bihar Animal Sciences University, Patna-800014, Bihar*
**Author for Correspondence: punam.gdmc@gmail.com*

ABSTRACT
DNA barcoding has emerged as a powerful tool for species identification and phylogenetic studies, particularly in taxonomically complex genera. The present study evaluates the effectiveness of the *matK* gene locus for species-level resolution among 17 *Aloe* species using an *in silico* approach. A total of 56 *matK* sequences corresponding to voucher specimens were curated and their 700 bp aligned for comparative analysis. Multiple sequence alignment and phylogenetic analyses were performed using MEGA 12 software employing the Maximum Likelihood method. The results revealed minimal intraspecific variation in most species, with 13 out of 17 showing no divergence among individuals. Interspecific divergence ranged from 0.000 to 0.017, with certain species pairs indicating close evolutionary relatedness. Phylogenetic analysis successfully resolved 10 out of 17 species into well-supported monophyletic clades, demonstrating over 58% species resolution efficiency. These findings underscore the potential of the *matK* gene as a reliable barcode for species identification within *Aloe*. However, the limited availability of high-quality sequences in public databases and the low variability among some species suggest the need for expanded sampling and possibly the use of multi-locus barcoding approaches for enhanced taxonomic resolution.

Keywords: *Aloe, matK, rbcL, DNA Barcoding, Phylogeny*

INTRODUCTION
The genus *Aloe*, belonging to the family Asphodelaceae, comprises over 500 species widely distributed across the arid and semi-arid regions of Africa, the Arabian Peninsula, and parts of Asia (Grace *et al*., 2011; Newton, 2001). These species are renowned for their medicinal, cosmetic, and economic importance, particularly *Aloe vera*, which has achieved global recognition for its therapeutic applications. Despite their high commercial and ethnobotanical value, the taxonomic identification of *Aloe* species remains a challenging endeavor due to significant morphological similarities, phenotypic plasticity, hybridization, and environmental influences (Grace *et al*., 2011). This complexity necessitates a more robust, accurate, and reproducible method for species identification beyond traditional morphological keys.

In recent decades, molecular tools have revolutionized the field of plant systematics and biodiversity assessment. One such tool that has emerged with significant promise is DNA barcoding (Hebert *et al*., 2003). DNA barcoding refers to the use of a standardized, short DNA sequence from a specific genome region to identify and differentiate species. The ideal barcode should be easily amplifiable, possess a high degree of interspecific variation, and maintain low intraspecific divergence. In plants, several loci from the chloroplast and nuclear genomes have been evaluated for their potential as DNA barcodes, with varying degrees of success. Among these, chloroplast-encoded markers such as *rbcL*, *matK*, *trnH-psbA*, and *atpF-atpH*, as well as nuclear markers like ITS, have been most extensively studied (Hollingsworth et al., 2009).

The maturase K (*matK*) gene, located within the intron of the *trnK* gene in the chloroplast genome, has emerged as one of the most promising candidates for plant DNA barcoding (CBOL Plant Working Group,

2009). *matK* is involved in group II intron splicing and is highly conserved across angiosperms. Despite its functional constraints, it exhibits relatively high substitution rates compared to other plastid genes, making it valuable for phylogenetic analysis and species-level discrimination (Dunning andSavolainen, 2010). The Consortium for the Barcode of Life (CBOL) Plant Working Group has recommended *matK*, along with *rbcL*, as a core DNA barcode for land plants (CBOL Plant Working Group, 2009). However, the performance of *matK* as a barcode varies among different plant groups, and its effectiveness in distinguishing closely related taxa, such as those in the *Aloe* genus, warrants closer investigation (Roy *et al*., 2010).

The application of *in silico* analysis in DNA barcoding offers a rapid, cost-effective, and scalable approach to evaluate the discriminatory power of various loci before engaging in laborious laboratory experiments (Kress and Erickson, 2008). *In silico* studies utilize publicly available sequence databases such as NCBI GenBank to retrieve barcode data, which can then be aligned, analyzed for genetic distances, and used to construct phylogenetic trees (Lahaye*et al*., 2008). This approach is particularly useful for preliminary assessments, primer evaluation, sequence divergence analysis, and detection of potential gaps in available barcode data. Moreover, it facilitates large-scale comparative studies across taxa, enabling the identification of universal or taxon-specific barcodes (Hollingsworth, 2011).

Despite the increasing volume of DNA sequence data in public repositories, comprehensive *in silico* evaluations of *matK* loci in *Aloe* species remain limited (Seberg and Petersen, 2009). Most existing studies focus on a handful of species or rely on multiple loci without isolating the performance of *matK*. Additionally, sequence quality, length variation, annotation discrepancies, and the presence of incomplete or chimeric sequences in public databases can further complicate barcode assessment (Roy *et al*., 2010). Therefore, a focused and methodologically sound *in silico* analysis of *matK* in the context of *Aloe* spp. is essential for understanding its potential and limitations as a standalone barcode.

This study aims to fill this gap by undertaking a systematic *in silico* evaluation of the *matK* locus across multiple *Aloe* species. By retrieving *matK* sequences from GenBank, aligning them using robust bioinformatics tools, and constructing phylogenetic trees, this work seeks to assess the interspecific variability, barcode gaps, and discriminatory capacity of the *matK* gene (Newmaster *et al*., 2006). The ultimate goal is to determine whether *matK* alone can serve as a reliable marker for species-level identification in *Aloe*, and if not, to highlight the need for complementary markers or multi-locus approaches (Hollingsworth, 2011).

In addition to its taxonomic implications, the outcomes of this study have broader applications in biodiversity conservation, sustainable utilization, and regulatory frameworks concerning *Aloe* species (Fay, 2011). With the increasing global trade in *Aloe*-based products, there is an urgent need for accurate molecular identification methods to authenticate species, detect adulterants, and ensure compliance with international biodiversity and trade agreements such as CITES. DNA barcoding, supported by *in silico* evaluations, can provide a scientific basis for such efforts, promoting both ecological stewardship and economic integrity (NCBI GenBank).

Furthermore, the methodological approach adopted in this study can serve as a template for similar evaluations in other plant genera where morphological ambiguity and taxonomic uncertainty prevail. It underscores the role of computational biology in enhancing traditional taxonomic practices, bridging the gap between molecular data and real-world applications (Tautz *et al*., 2003; DeSalle *et al*., 2005). By critically assessing the *matK* locus in a computational framework, this research contributes to the growing body of literature advocating for integrative, data-driven solutions in plant systematics and molecular ecology.

In conclusion, the integration of DNA barcoding and *in silico* methodologies offers a powerful toolset for resolving taxonomic complexities within the *Aloe* genus. This paper, titled "An *In Silico* Approach for Evaluation of *matK* Loci for DNA Barcoding of *Aloe* spp.", seeks to investigate the efficacy of *matK* as a single-locus barcode through comprehensive sequence analysis and phylogenetic evaluation. It aims to provide insights into the molecular diversity of *Aloe* species and support the development of robust

identification systems for conservation, trade, and research purposes (Sharma *et al*., 2020; Hollingsworth and Hollingsworth, 2010).

MATERIALS AND METHODS

DNA sequences of the *matK* gene corresponding to voucher specimens of *Aloe* species were included in this study. A total of 17 *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 (Suesatpanit *et al*., 2017). The selected sequences were manually curated to ensure a uniform length of 700 base pairs (positions 543–1242) 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 genetic distance estimation and phylogenetic inference, were also conducted using the same

Table 1: List of *Aloe* species under investigationalong with their sequence details

Aloe Species	Gene bank Accession number of the sequences used in the study	No of *matk* gene sequence selected for analysis	No of *matk* gene sequence downloaded
Aloe vera	GQ434051.1, JN228939.1, KP072730.1 KP072729.1, KP072728.1, KP072727.1 KP072726.1, KP072725.1, KC893719.1 GQ434050.1, GQ434049.1, GQ434048.1 KY556640.1 GQ434047.1	14	25
Aloe maculata	KP072720.1, KP072721.1, KP072722.1 JX517325.1, JQ412193.1	5	09
Aloe brevifolia	JQ024117.1, JX517854.1	2	04
Aloe castanea	KC893700.1, JQ024120.1	2	02
Aloe comosa	JQ024123.1, JQ024124.1	2	03
Aloe glauca	JQ024133.1, JQ024134.1	2	05
Aloe lineata	JQ024146.1, JQ024147.1, JQ024148.1 JQ024149.1	4	06
Aloe marlothii	KC893708.1, JF270641.1	2	02
Aloe melanacantha	JX517575.1, JQ024150.1	2	04
Aloe microstigma	JQ024151.1, JQ024152.1, JQ024153.1	3	04
Aloe nyeriensis	JQ435526.1, KU748254.1, KU748259.1	3	07
Aloe pearsonii	JQ024154.1, KC893709.1	2	03
Aloe perfoliata	JQ024155.1, JQ024156.1, JX517501.1	3	03
Aloe powysiorum	KU748251.1, KU748260.1, KU748291.1	3	03
Aloe purpurea	KX270416.1, KX270418.1	2	03
Aloe spicata	KC893712.1, JF270642.1	2	02
Aloe striata	KC893713.1, JQ024162.1, JQ024161.1	3	04

software. Phylogenetic analysis was performed using the Maximum Likelihood (ML) 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).

RESULTS AND DISCUSSION

The search using key word *Aloe+matk* yielded multiple sequences across different species mentioned in Table 1. A total of 56 sequences belonging to 17 *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 17 species and 56 curated sequences. The length of sequence available in Genebank varied from as short as 262 in *A gluca* to full length 1557 bases *matk* gene in *Aloe vera*. The selection of 700 sequence length of *matk* genes was aimed to get result with accuracy. Notably, *Aloe vera* had the highest representation with 25 sequences available, of which 14 were selected for analysis.

***Estimation of sequence divergence*:** The variation among 17 *Aloe* species under study as well as within individuals of a species was determined and presented in table as evolutionary divergence (p distance value).Out of all species examined, 13 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 four individuals ranged between 0.009 to 0.015 in

Table:2 Average evolutionary divergence of *matK* regions

Aloe vera	0	0
Aloe maculata	0.009	0.002
Aloe brevifolia	0	0
Aloe castanea	0	0
Aloe comosa	0	0
Aloe glauca	0.015	0.005
Aloe lineata	0	0
Aloe marlothii	0	0
Aloe melanacantha	0	0
Aloe microstigma	0.010	0.003
Aloe nyeriensis	0	0
Aloe pearsonii	0	0
Aloe perfoliata	0.003	0.001
Aloe powysiorum	0	0
Aloe purpurea	0	0
Aloe spicata	0	0
Aloe striata	0	0

the species *A. maculata, A. gluca, A. microstigma* and *A. powsoyrium* which is comparatively lesser in comparison to similar type of study (Ho and Nguyen, 2020). The aforesaid finding suggests that the *matk* gene is highly conserved *Aloe*s pecies and can be of taxonomical importance. The *matk* gene based

variability among *Aloe* species varied from 0.000 to 0.017 (Tables 2) with no interspecies variation between *A. castenca* and *A comosa* which reveals presence of uniformity in evolution between them.

***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 10 out of 17 *Aloe* species, includes *Aloe vera, A. lineata, A. marlothii, A. melanacantha, A. nyeriensis, A. pearsonii, A. powysiorum, A. spicata, A. purpurea* and *A. striata* based on

Table:3 Evolutionary divergences in *matk* sequence pairs between 17 species of *Aloe* genus

Aloe vera																
Aloe maculata	0.010															
Aloe brevifolia	0.006	0.008														
Aloe castanea	0.011	0.009	0.009													
Aloe comosa	0.011	0.009	0.009	0.000												
Aloe glauca	0.012	0.012	0.008	0.008	0.008											
Aloe lineata	0.013	0.010	0.010	0.001	0.001	0.008										
Aloe marlothii	0.004	0.011	0.007	0.013	0.013	0.014	0.014									
Aloe melanacantha	0.007	0.010	0.004	0.010	0.010	0.011	0.011	0.009								
Aloe microstigma	0.012	0.012	0.011	0.007	0.007	0.009	0.007	0.013	0.012							
Aloe nyeriensis	0.004	0.011	0.007	0.013	0.013	0.014	0.014	0.006	0.009	0.013						
Aloe pearsonii	0.007	0.008	0.004	0.010	0.010	0.011	0.011	0.009	0.003	0.012	0.009					
Aloe perfoliata	0.007	0.009	0.003	0.010	0.010	0.009	0.011	0.008	0.003	0.012	0.008	0.003				
Aloe powysiorum	0.004	0.011	0.007	0.013	0.013	0.014	0.014	0.006	0.009	0.013	0.000	0.009	0.008			
Aloe purpurea	0.004	0.009	0.004	0.010	0.010	0.011	0.011	0.006	0.006	0.011	0.006	0.006	0.005	0.006		
Aloe spicata	0.006	0.008	0.003	0.009	0.009	0.009	0.010	0.007	0.004	0.011	0.007	0.004	0.004	0.007	0.004	
Aloe striata	0.011	0.009	0.009	0.014	0.014	0.015	0.016	0.013	0.010	0.017	0.013	0.007	0.010	0.013	0.010	0.009

the phylogenetic tree constructed (Figure 1). The bootstrap values below 50% have been excluded from the figure for clarity. The study indicated that *matk* gene has more than 58% species resolution ability for *Aloe* species, However 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 similar type of finding was obtained in a study of 124 DNA sequences including ITS, *matK* and *rbcL* DNA barcode loci on 11 orchid species, which revealed that the discrimination capacity of *matK* and ITS loci has more potential for genetic classification at its genus and species level (Ho V T, 2021). 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 (Singh *et al.*2016). The present study includes a relatively high sample size of 14 sequences for *Aloe vera*, compared to four for *Aloe lineata*, and three each for *A. powysiorum, A. striata,* and *A. nyeriensis*, while the remaining species are represented by two samples each. Despite this variation in sample size, the *matK* gene-based analysis demonstrated strong potential in resolving species-level relationships. However, the availability of *matK* 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).

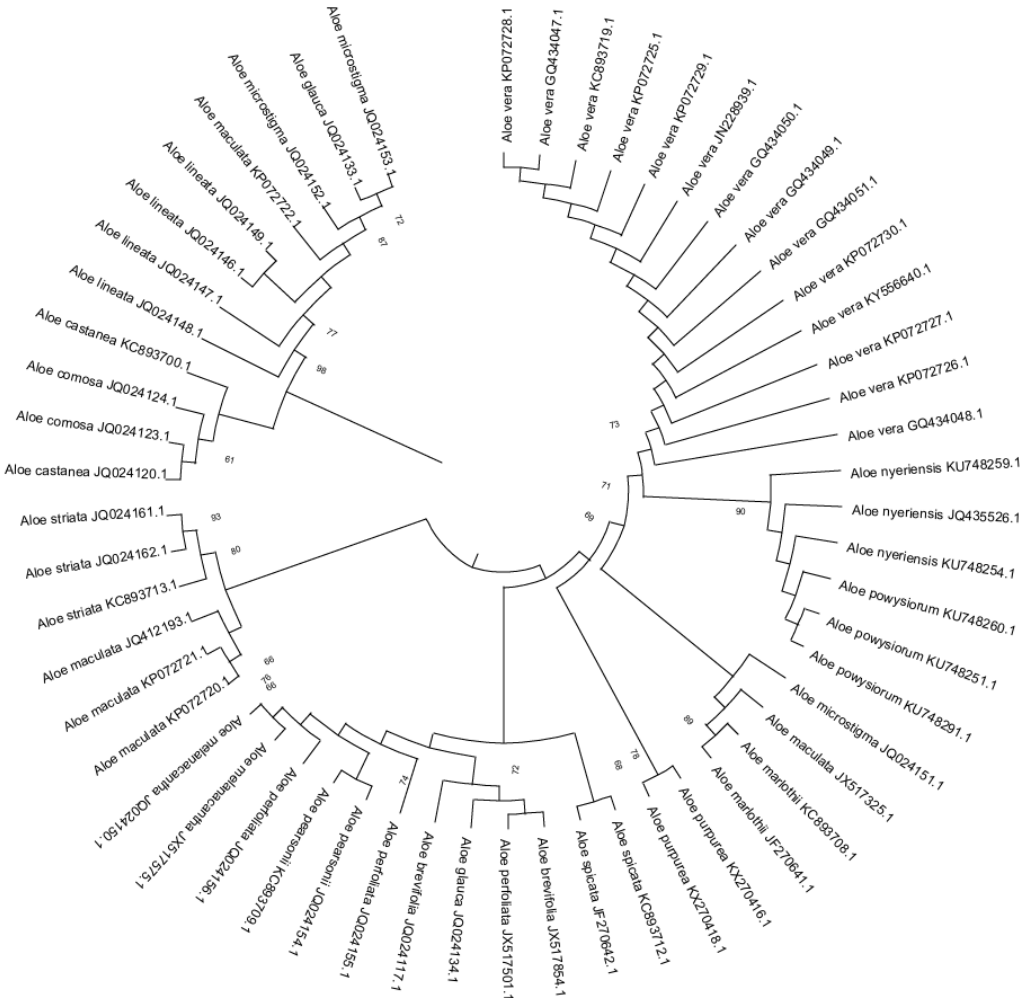

Figure 1: Maximum Likelihood (ML) tree with 1000 bootstrap replicated based on *matk* sequences

CONCLUSION

The present study highlights the potential of the *matK* gene as an effective DNA barcode for species-level identification within the genus *Aloe*. Through the analysis of 56 curated sequences representing 17 species, the *matK*-based phylogenetic approach successfully resolved 11 species into distinct monophyletic clades, indicating a resolution efficiency of over 58%. Despite the high conservation observed in some species, including *Aloe vera*, the gene still exhibited sufficient interspecific variability for effective discrimination in most cases. However, the study also reveals certain limitations, including the sparse availability of *matK* sequences for some species in public databases and limited sample sizes, which may affect resolution accuracy. The complete absence of variation between some closely related species suggests the need for complementary loci (such as *rbcL* or ITS) to enhance species delimitation. 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 acknowledge Prof. Rimjhim Sheel, Prof. In Charge, Ganga Devi Mahila Mahavidyalaya for providing conducive research environment and invaluable scientific inputs and Dr. Bhavya Jha, Assistant Professor, Department of Zoology, GDMM, Patliputra University, Patna for scientifically fruitful discussions.

REFERENCES

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW (2012). GenBank. *Nucleic acids research.***41**(D1) D36-42.https://doi.org/10.1093/nar/21.13.2963

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

DeSalle R, Egan MG and Siddall M (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1462), 1905–1916. https://doi.org/10.1098/rstb.2005.1722

Dunning LT and Savolainen V (2010). Broad-scale amplification of *matK* for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society*, **164**(1), 1–9. https://doi.org/10.1111/j.1095-8339.2010.01071.x

Fay MF (2011). Orchid conservation: How can we meet the challenges in the twenty-first century? *Botanical Studies*, **52**, 367–374.https://doi.org/10.1186/s40529-018-0232-z

Grace OM, Simmonds MSJ, Smith, GF, van Wyk AE, Klopper RR, Buerki Sand Ronsted N (2011). Documented utility and potential of the *Aloe* genus: A review. *South African Journal of Botany*, **77**(4), 980–987. https://doi.org/10.1007/s12231-009-9082-7

Hebert PDN, Cywinska A, Ball SL andde Waard, JR. (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

Hollingsworth ML and Hollingsworth PM (2010). DNA barcoding: A molecular tool for species identification. *Biologist*, **57**(4), 150–155.

Hollingsworth PM (2011). Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, **108**(49), 19451–19452. https://doi.org/10.1073/pnas.1116812108

Ho VT (2021) Evaluating the Effectiveness of Three DNA Bar Code Loci to Classify Jewel Orchids Using *In silico* Approach bioscience biotechnology research communications **14** (2) http://dx.doi.org/10.21786/bbrc/14.2.23

Ho VT, Nguyen PM (2020). An *in silico* approach for evaluation of rbcl and *matk* loci for DNA barcoding of cucurbitaceae family, Biodiversitas 21:08, 3879-3885https://doi.org/10.13057/biodiv/d210858

Kress WJ and Erickson DL (2008). DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences*, **105**(8), 2761–2762. https://doi.org/10.1073/pnas.0800476105
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
Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, ... & Savolainen V (2008). DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences*, **105**(8), 2923–2928 https://doi.org/10.1073/pnas.0709936105
Newmaster SG, Fazekas AJ and Ragupathy S (2006). DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. *Canadian Journal of Botany*, **84**(3), 335–341 https://doi.org/10.1139/b06-047
Newton LE (2001). *Aloe* taxonomy: the current status of the genus. *British Cactus & Succulent Journal*, **19**, 11–17.
Roy S, Tyagi A, Shukla V, Kumar A, Singh U M, Chaudhary LB ... & Sharma V (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. Jul 10;18(4):1105-15.
Seberg O and Petersen G (2009). How many loci does it take to DNA barcode a crocus? *PLoS ONE*, **4**(2), e4598 https://doi.org/10.1371/journal.pone.0004598
Sharma J, Devi R, Choudhary MK and Rana JC (2020). DNA barcoding for authentication and conservation of plant genetic resources: A review. *Indian Journal of Agricultural Sciences*, **90**(11), 2063–2070.
Sikdar S, Tiwari S, Sapre S and Thakur VV (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.
Singh A, Drishti NG and Mohanty A. (2016) *In silico* analysis of sequence variation in rbcL gene to assess phylogenetic relations in Setaria species. *In*: International Conference on Innovative Research in Agriculture, Food Science, Forestry, Horticulture, Aquaculture, Animal Sciences, Biodiversity, Ecological Sciences and Climate Change (AFHABEC-2016) Oct (Vol. 22).
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
Tautz D, Arctander P, Minelli A, Thomas RH, &Vogler AP (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, **18**(2), 70–74. https://doi.org/10.1016/S0169-5347(02)00041-1