Invasiveness Identification: A Study Case from *Lantana*

Inggar Damayanti[1*], Muhammad Rifqi Hariri[2], Melza Mulyani[2]

[1]Forestry Department, Faculty of Agriculture, University of Lampung, Jl. Prof Soemantri Brojonegoro No.1 Gedung Meneng Kedaton Bandar Lampung 35144

[2]Research Center for Plant Conservation and Botanic Gardens – Bogor Botanic Gardens, Indonesian Institute of Sciences, Jl. Ir. H. Juanda No. 13 Bogor

** Corresponding author: Inggar.damayanti@fp.unila.ac.id*

ABSTRACT

Invasion by invasive species represents one of the greatest threats to biodiversity worldwide, causing degradation and loss of habitat. Among them, one species belonged to the Verbenaceae family, namely Lantana camara, which includes 100 of the world's worst invasive species. Distinguishing invasive from non-invasive species based on morphology alone is often difficult for plants in a vegetative stage, especially in Lantana, where they have complex morphological characters. In this regard, DNA barcoding may become a good alternative. This study aimed to select and provide a DNA barcode region that capable of distinguishing the invasive and non-invasive Lantana. Four DNA Barcode markers available in the sequence database (NCBI and BOLD), namely matK, rbcL, psba-trnh, and ITS2, were used to identify the invasiveness of various Lantana. A total of 132 data sequences from 17 species of Lantana were collected. The sequences were aligned and constructed into a dendrogram using MEGA X through the Neighbor-Joining method. This study shows that it is possible to distinguish Lantana camara from a series of closely related congeners by plastid base gene (matk and rbcl). The constructed phylogeny tree shows that invasive species Lantana camara was in a different clade with non-invasive Lantana.

***Keywords**: DNA Barcode, Lantana, invasive alien plant species*
Diterima, 28 Juni 2021
Disetujui, 16 Desember 2021
Online, 30 Desember 2021

INTRODUCTION

The presence of invasive species is one of the many threats to biodiversity throughout the world. It is affecting ecosystems and contributing to the local extinction of native species, invasive species can also cause damage to the socio-economic sector (Pimentel *et al.*, 2005). Increasing the spread of invasive species is a result of increased transportation and trade. Many invasive species have entered new areas through commerce, either purposely, such as garden or aquarium plants, or by accident as stowaways or weeds (Ghahramanzadeh, 2013). Among all the noxious invasive species, one species belonged to the Verbenaceae family, namely *Lantana camara*, which includes 100 of the world's worst invasive species. The genus *Lantana* contained over 150 species (Chen & Gilbert, 1994; Atkins, 2004). *Lantana* is mainly native to subtropical

and tropical America, but a few taxa are indigenous to tropical Asia and Africa (Ghisalberti, 2000). The *L. camara*, commonly known as wild or red sage, is the most widespread species of the *Lantana* genus. Apart from its popularity as a garden plant, *L. camara* is an aggressive, obligate, and outbreeder weed that has invaded vast expanses of pastures, orchards, and forest areas in many tropical and subtropical regions (Parsons, 1992).

One way of preventing the spread of known invasive species would be to ban their import. We must be able to unequivocally distinguish them from related, non-invasive species to make this feasible. This can be a problem due to the difficulty of distinguishing invasive species from non-invasive species based on plant morphology in a vegetative stage. Moreover, the genus of *Lantana* is challenging to classify since the species are not stable and hybridization is widespread, the shape of inflorescence changes with age, and flower colors vary with age and maturity (Ghisalberti, 2000).

DNA barcoding is an alternative method to identify species through a short and standardized DNA region, called DNA barcode, across all possible forms of life (Hebert, 2003). The selection of plant barcode loci involved a complicated compromise between universality and discrimination. The ideal barcode loci would require a certain level of variation for discriminatory power. However, they also must be somewhat conservative for universality and ease of alignment (Zhang, 2013). In principle, DNA barcodes contain variation that can be posed as a character to differentiate species (Amandita, 2018). The number of candidate gene regions was suggested as potential barcodes for plants, including coding genes and noncoding genes in the nuclear and plastid genomes (Chase *et al.*, 2007; Kress & Erickson, 2007). DNA barcodes in plants generally use *rbc*L and *mat*K markers. Besides that, *trn*H-*psb*A also be a suitable marker to discriminate among closely related species (Kress & Erickson, 2007). Moreover, nuclear genomic regions, such as the internal transcribed spacer (ITS) region, were also suggested as potential DNA barcodes due to high levels of interspecific sequence variability (Kress *et al.*, 2005, Cowan & Fay, 2012).

Bio-monitoring of invasive species is one applied field that urgently needs the DNA barcode technique (Darling, 2007) because the DNA characters have relatively more consistent properties than morphological characters. Early estimation of the best DNA barcoding primer selection for invasive species identification needs to be done. The investigation of these markers will contribute to the development of helpful barcode information for invasive plant identification. This study aimed to select and provide a DNA barcode region capable of distinguishing the invasive and non-invasive *Lantana* using DNA Barcode markers available in the sequence database of NCBI and BOLD, namely *mat*K *rbc*L, *psb*A-*trn*H, and ITS2.

MATERIAL AND METHOD

Materials used in this study were sequence data collection of *Lantana* genus from four DNA Barcode markers, namely *mat*K, *rbc*L, *psb*A-*trn*H, and ITS2 generated from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/genbank/) and The Barcode of Life Data System (BOLD) (http://boldsystems.org/). The procedure of sequence data collection is called the data mining method. Details of the DNA sequences used in this study are shown in Table 1.

Table 1. DNA sequence extracted from NCBI and BOLD

No	Species	Accession Number
1	*L. angustifolia*	HM120857
2	*L. camara*	GENG277-14, GU135140, JF265499, JQ594382, JQ594383, JQ594384, JQ594385 ,JQ594386, JQ594387, JQ594388, JQ594389, JQ618495, JQ618496, JQ618497, JQ618498, JX571858, KF425765, KP208916, KU556643, KU569183, KX78391, KY627498, MF694736, MF694990, MG784921, MH050106, MH050107, MH050108, MH050109, MH549895, MHPAC1389-11, MHPAC1390-11, MHPAC1391-11, PPBI007-16, PPBI023-16, PPBI024-16, SDH2086-14, MK290473, MH621559, MH552322, MG784975, MF694861, KX783702, JX495729, JQ589773, JQ589442, JQ589441, JQ589440, JQ589439, JQ589438, JQ589437, JQ589436, JQ589435, JQ589434, JF270846, HM853859, HM850972, GU134977, GQ429057, AF315303, MK260675, MH621960, KU198271, JQ618444, JQ618443, JQ618442, GU135307, GQ429115, PPBI024-16, PPBI023-16, MHPAD988-09, PPBI007-16, MHPAD987-09, MHPAD986-09, MHPAD1051-09, MG730661, MG730660, MG730659, MG730658, MG256271, KY700391, KY700390, KY700389, KX115485, KX115484
3	*L. canescens*	MH549896, MH621579, HM853857, MH621961
4	*L. depressa*	KJ773614, MH549897, MH621558, KJ772886, MH621962, FJ004801
5	*L. hirsuta*	HM120856
6	*L. hirta*	HG963495
7	*L. hodgei*	HM120851
8	*L. horrida*	DQ463783, HM120852
9	*L. involucrata*	KJ082380, MH549898, MH621573, KJ012653, MH621963
10	*L. micrantha*	HM120854
11	*L. montevidensis*	SDH2087-14.1
12	*L. rugosa*	JF265500, JX572712, JX517746, JF270847
13	*L. salvifolia*	GENG709-14
14	*L. scabrida*	HM120860, HM120859, HM120858
15	*L. strigocamara*	HM120861, HM120853
16	*L. trifolia*	JQ594390, JQ594391, JQ594392, MHPAC1395-11, MHPAC1396-11, MHPAC1397-11, JQ589445, JQ589444, JQ589443, MHPAD1059-09, MHPAC1395-11, MHPAC1397-11, MHPAC1396-11
17	*L. urticifolia*	HG963500

Data analysis was carried out using sequence data collection of the *Lantana* genus, which was obtained from data mining. Each of these sequence barcodes was assigned to a particular taxon by comparing it with the nucleotide sequences in Gen Bank database NCBI using Basic Local Alignment Search Tools (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Moreover, the results of sequence identification were cross-checked with the morphological identification results from the sample identity.

The cross-checked results between morphological and molecular identification were counted into three levels: species, genus, and family. The following decisions were made to identify marker efficiency.

Sequence alignment was performed by using the ClustalW program (Thompson, Higgins, & Gibson, 1994) embedded in MEGA X (Tamura et al., 2013) for each marker. The alignment results were subsequently checked for ambiguities caused by the presence of indels and/or substitutions and edited if necessary. Based on the aligned sequences, phylogenetic trees were reconstructed using MEGA X (Tamura et al., 2013) with neighbor-joining (NJ) algorithms.

RESULT AND DISCUSSION

DNA barcoding can be an effective tool to identify plant or animal species faster, does not require a complete sample, and does not require special skills. The research to determine the effectiveness of gene loci as DNA barcodes in plants has been widely carried out (Kress et al., 2005). This study uses data available on the Genebank (NCBI) to obtain barcode loci that can be used to identify species of the *Lantana* genus. This is the first step to provide initial information to prevent the spread of species from other *Lantana* genera as an effort to eradicate invasive species.

The results of sequence data mining from GenBank using *mat*K, *rbc*L *psb*A-*trn*H, and ITS2 molecular markers obtained a total of 132 accessions labeled with 17 different species names. The *rbcL* universality as DNA barcode was observed in this study as the highest amount of sequence data collected. A total number of 52 accessions from 8 species was obtained. We also obtained 34 accessions from 6 species for *mat*K marker, 33 accessions from 10 species for ITS2, and 13 accessions from 6 species for *psb*A-*trn*H marker. Among all the sequence data, *L. camara* species are the most dominant. The 85 accessions from a total of 132 accessions belonging to the species of *L. camara*. The results of the aligning sequences showed that the average base pair of *mat*K were 763 bp, 714 bp for *rbc*L, 340 bp for *psb*A-*trn*H, and 524 for ITS2. The distribution of sequence data for each species is shown in Figure 1.

e-ISSN 2621-4113
p-ISSN 2615-8353

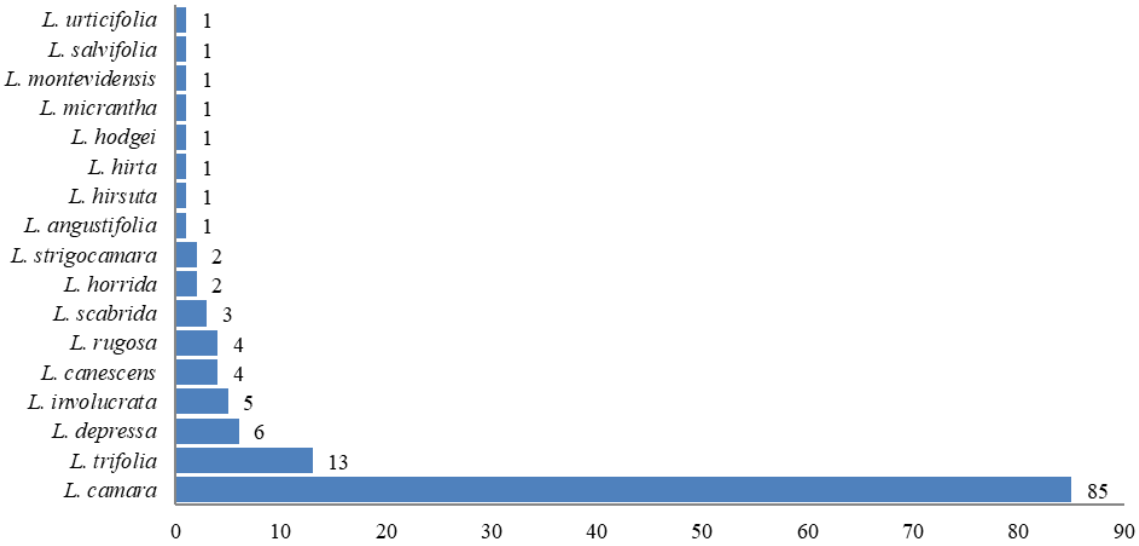

Figure 1. The composition of *Lantana* accession of sequence collected from NCBI

As one way to evaluate the success rate of species identification, we compared the results from morphological identification with the results from molecular identification using Basic Local Alignment Search Tools (BLAST-n). Among all of the markers, the highest match between morphological and molecular identification was at the species level, with an identity value of 100%. The identity value is the similar percentage of the DNA sequence inputted with the DNA sequence in GenBank. High identity values indicate a high nucleotide sequence match. In this study, the matched identification at species level was higher with *mat*K and *rbc*L with an identity value of 100% for both markers compared to *psb*A-*trn*H and ITS2 (99,89% and 99,94%, respectively).

According to BLAST results, *mat*K has higher overall species identification success as all the samples can be identified (100%), followed by *psb*A-*trn*H, ITS2, and *rbcL* (92,31%, 87,88%, 84,62%, respectively). The mismatch between morphological identification and DNA identification results could be due to several reasons. A specimen could be misidentified when it was found or could also have the highest similarity to a reference sequence that was falsely identified through morphological characters. Another factor affecting species identification success using DNA barcoding is the availability of nucleotide data of the corresponding taxa in the DNA sequences database such as GenBank and BOLD (Amandita, 2018). An accurate and complete molecular database, especially for plant species, will hopefully be developed in the future.

Phylogenetic trees were constructed based on multiple sequence alignments of *mat*K, *rbc*L *psb*A-*trn*H, and ITS2 using the Kimura 2-parameter method. This method uses transitional and transversion parameters to measure the percentage difference in genetic distance between samples (Nei & Kumar, 2000). The method used is Neighbor-Joining (NJ) with a bootstrap value of 1000x. This method effectively counts the nucleic

acid differences used in identifying the species and assesses its similarity. *Lantana* species are clustered together for the four primers and separated clearly with its outgroup, *Aloysia virgata*.

The tree generated from *rbc*L showed high similarity sequences of those *Lantana* specimens compared to the other primers (Fig. 2). The tree generated from *mat*K (Fig. 3) and *psb*A-*trn*H (Fig. 4) showed a clear and distinct clade consisted of *L. camara* and *L. depressa* with the other species. On the other hand, though ITS2 can distinguish *A. virgata* and *Lantana* species, this primer seemed to have a pretty slight capability in identifying and determining those *Lantana* (Fig. 5).

The universality owned by *the rbc*L sequence proved that this sequence tends to be conserved and has a low rate of nucleotide mutation. Meanwhile, *mat*K and *psb*A-*trn*H sequences have a higher chance of mutation that can be used to distinguish *Lantana* species. However, the ITS2 sequence showed a higher rate of mutation that is difficult to use, except for *L. trifolia. Lantana* cannot be identified because they have high similarities and varieties in their vegetative organs (Silva, 1999; Salimena, 2002). As stated above, the recommended DNA barcoding regions can provide information and help precise species identification to support morphological identification, which is difficult when floral organs are not found.

e-ISSN 2621-4113
p-ISSN 2615-8353

Figure 2. Phylogenetic tree generated from *rbc*L sequence

e-ISSN 2621-4113
p-ISSN 2615-8353

Figure 3. Phylogenetic tree generated from *mat*K sequence

e-ISSN 2621-4113
p-ISSN 2615-8353

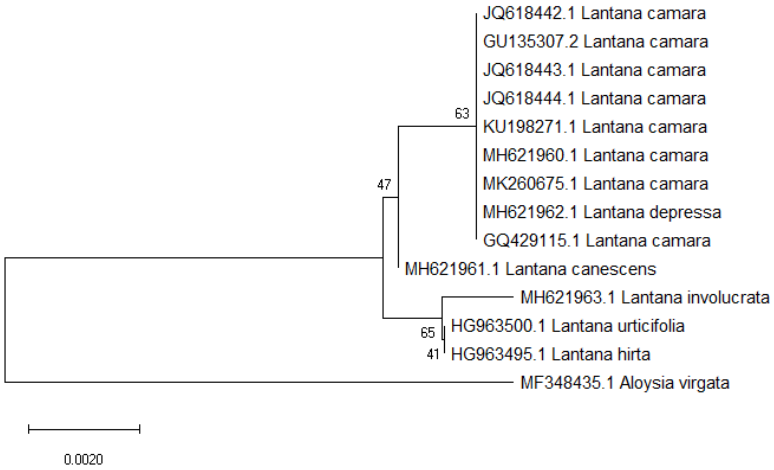

Figure 4. Phylogenetic tree generated from *psb*A-*trn*H sequence

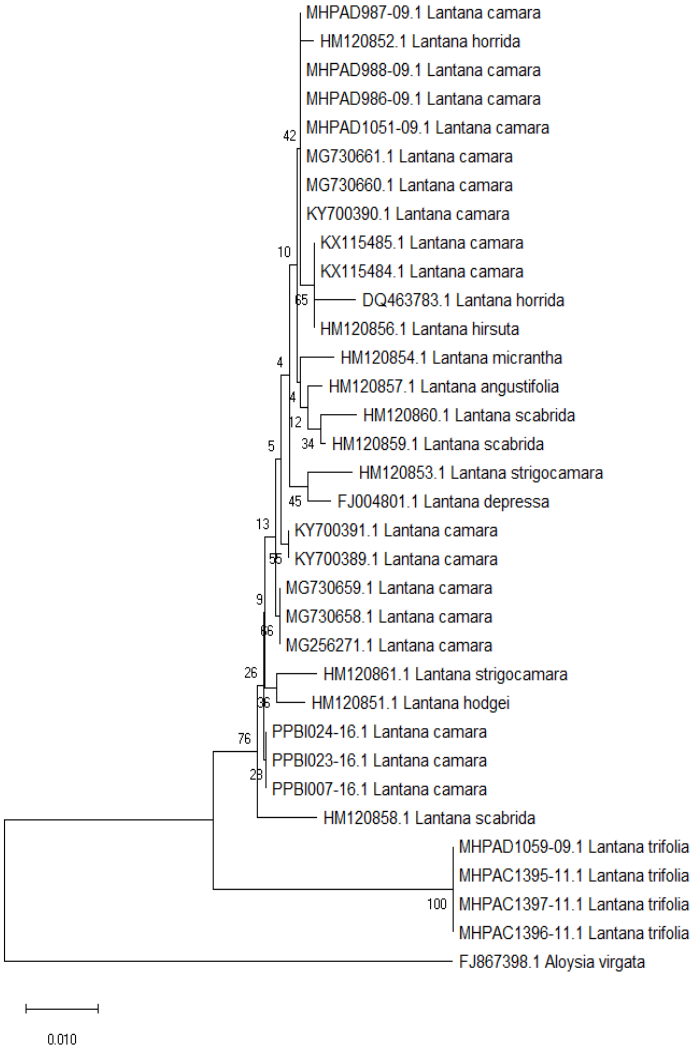

Figure 5. Phylogenetic tree generated from ITS sequence

CONCLUSION

Among the DNA Barcode regions used to construct the phylogenetic tree in this study, the *rbc*L and ITS2 are regions that cannot distinguish each *Lantana* species compared to the *mat*K and *psb*A-*trn*H region. The *mat*K showed higher reliability in determining *Lantana* species than *psb*A-*trn*H. Therefore, we recommend the *mat*K region as a barcode marker for *Lantana* species to distinguish the invasive *Lantana* its ornamental kind.

REFERENCES

Agaldo, J. A. 2020. Ecology of *Lantana* camara: An invasive plant species in Amurum Forest Reserve, Jos Plateau, Central Nigeria. *African Journal of Ecology*. 58(2), 291-298.

Amandita FY, Rembold K, Vornam B, Rahayu S, Siregar IZ, Kreft H, Finkeldey R. 2019. DNA barcoding of flowering plants in Sumatra, Indonesia. *Ecology and Evolution*. 9:1858-1868

Atkins, S. 2004. *Verbenaceae. Flowering Plants. Dicotyledons*. Berlin. Springer.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Wilkinson M. 2007. A proposal for a standardized protocol to barcode all land plants. *Taxon*. 56:295–299.

Chen G, Gilbert MG. 1994. Verbenaceae. In: Wu ZY & Raven PH (eds.). Flora of China. Vol. 17. Science Press. Beijing.

Cowan RS, Fay MF. 2012. Challenges in the DNA barcoding of plant material. *Molecular Biology*. 862: 23–33.

Darling JA, Michael AE, Blum J. 2007. DNA-based methods for monitoring invasive species: a review and prospectus. *Biol Invasions*. 9:751–765

Ghahramanzadeh R, Esselink G, Kodde LP, Duistermaat H, Van Valkenburg JLCH, Marashi SH, Smulders MJM, Van De Wiel CCM. 2013. Efficient distinction of invasive aquatic plant species from non-invasive related species using DNA barcoding. *Molecular Ecology Resources*. 13:21–31.

Ghisalberti EL. 2000. *Lantana* camara L. Verbenaceae. *Fitoterapia*. 71:467-486

Hajibabaei M, de Waard JR, Ivanova NV, Ratnasingham S, Dooph RT, Kirk SL, Hebert PDN. 2005. Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B*. 360: 1959–1967.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*. 270:313–321.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trn*H-*psb*A spacer region. *PLoS ONE*. 2:e508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*. 102: 8369–8374.

Nei M, Kumar S. 2000 Molecular evolution and phylogenetics. New York. Oxford University Press.

Parsons WT, Cuthbertson EG. 1992. Noxious weeds of Australia. Melbourne. Inkata Press.

Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*. 52:273–288.

Salimena FRG. 2002. Novos sinônimos e tipificações em Lippia sect. Rhodolippia (Verbenaceae). *Darwiniana*. 121-125.

Shackleton RT, Witt AB, Aool W, Pratt CF. 2017. Distribution of the invasive alien weed, *Lantana camara*, and its ecological and livelihood impacts in eastern Africa. *African Journal of Range & Forage Science*. 34(1):1-11.

Silva TRS. 1999. Redelimitação e revisão taxonômica do gênero *Lantana* L. (Verbenaceae) no Brasil. Dissertation, PhD thesis, Universidade Federal de Sao Paulo.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution.* 30: 2725–2729.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL-W – Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Research*. 22: 4673–4680.

Zhang W, Fan X, Zhu S, Zhao H, Fu L. 2013. Species-specific identification from incomplete sampling: applying DNA barcodes to monitoring invasive solanum plants. *PLoS ONE.* 8(2):e55927.